Talking to AI-Equipped Investors: A Field Experiment

Yupeng Lin*
National University of Singapore
Zhenghui Ni
Renmin University of China
Yichen Su
Renmin University of China
Nianhang Xu
Renmin University of China

Abstract

We run a large-scale randomized controlled trial (RCT) with 2,722 Chinese listed firms to examine how perceived AI bots shape managers' disclosures and market reactions. In online earnings conferences (OECs), we randomize the questioner's identity as either human or AI while holding information content fixed. Managers are 13% less likely to answer AI-styled questions, with stronger effects when an image signals AI identity. Linking manager and firm traits to responses shows smaller declines at state-owned enterprises and larger declines for managers who are female, older, or have a computer science background. Conditional on responding, managers provide longer, timelier answers with more neutral narratives to AI-styled questions. Using high-frequency data, we find that investors withhold trading during AI-question windows. Overall, our results are consistent with AI playing a disciplinary role on voluntary disclosure.

Keywords: Randomized Controlled Trial (RCT), AI Bots, Information Processing, Voluntary Managerial Disclosure, Market Response
JEL: C9, D8, G14, G3.

* We thank Sean Cao, Fuxiu Jiang, Wei Jiang, Huailu Li, Shu Lin, Yibin Liu, Pinchuan Ong, David Reeb, Honglin Ren, Johan Sulaeman, Hun Tong Tan, Tu Xu, Wenjie Xue, Jingda Yan, Bernard Yin Yeung, Keyuan Zhang, Liandong Zhang, Weining Zhang, Zilong Zhang, Lingling Zheng, Luo Zuo, and seminar participants at Nanjing University, National University of Singapore, Renmin University of China, Shanghai Jiao Tong University, and Shanghai University of Finance and Economics, and South China University of Technology for their helpful suggestions. This research is registered with the AEA and is supported by IRB approval from Renmin Business School, Renmin University of China (2025R15). Yupeng acknowledges financial support from the National University of Singapore. Zhenghui acknowledges financial support from Renmin University of China and the National Natural Science Foundation of China (Grant no. 72502237). Nianhang acknowledges financial support from the National Natural Science Foundation of China (Grant no. 72225005 and 72495154).

1. Introduction

Generative AI is a new general-purpose information technology that lowers the cost of acquiring and interpreting corporate narratives, thereby expanding "information democracy" in capital markets (Dong et al. 2016; Blankespoor et al. 2020; Costello et al 2020; Cao et al. 2024; Chang et al. 2025; Bradshaw et al. 2025; Bertomeu et al. 2025a, 2025b). As AI becomes widely accessible, investors' information-processing capacity rises, potentially shifting disclosure incentives and market information aggregation to a new equilibrium (Bertomeu et al. 2025b). Yet we have little causal evidence on how firms respond to this change.[1]

In this paper, we provide such evidence using a randomized controlled trial (RCT) that exogenously changes managers' perceptions about the presence of AI-equipped investors. This intervention isolates a central margin: how managers adjust *voluntary disclosure*, that is, whether they communicate at all, and how they structure and emphasize narrative content when the effective audience includes AI bots. We emphasize that our design isolates the perception channel—how managers react to believing that the audience includes AI—and is not intended to capture the full effect of investors' AI use, which may also operate through changes in the content and precision of the questions themselves. In this sense, our estimates speak to one specific, cleanly identified margin rather than the aggregate impact of AI on disclosure.

It is challenging to identify how AI bots change voluntary disclosure in corporate communication. Managerial beliefs about audience sophistication are central to disclosure incentives yet empirically unobservable (e.g., Gao 2008; Gao and Liang 2013; Liang and Zhang 2019). Essentially, neither the extent of investor AI adoption nor managers' perceptions of it are

[1] An exception is Cao et al. (2023), who rely on machine downloads to show how firms cater to machine readers by making formatting of regulatory filings (i.e., 10-K and 10-Q filings) more machine-friendly. This paper seeks to focus on how firms adjust their mandatory disclosures in response to potentially rising machine readership.

directly observable.[2] Additionally, investor information needs covary with investor endowments and technological capacity, confounding interpretations of association. We address these issues by running a field experiment and manipulating only managers' perceptions of AI readership, following the logic of belief-based interventions (Alesina et al. 2018), while holding constant the information demand by investors.[3] Using this experiment, we investigate how firms adjust voluntary disclosure in response to perceived AI readership. We also provide suggestive evidence on how the market interprets those disclosures by drawing on intraday trading data.

Our setting is China's *online earnings conferences* (OECs): text-based, regulator-endorsed Q&A forums in which investors pose real-time questions to managers on public chat platforms. Relative to traditional earnings calls, OECs enable direct, unfiltered interaction and create reputational pressure to respond, making non-response an informative proxy for withholding. However, the absence of phrasing rules also leave flexibility for managers to provide a strategic narrative (Huang et al. 2014).

A core facet of our experiment design is to hold information quantity/content demand fixed, allowing us to isolate the effect of perceived AI readership on disclosure, beyond the demand-driven disclosure channel discussed in Cao et al. (2023). By varying only the perceived identity of the questioner, we causally shift managerial beliefs without changing the underlying informational content, thereby addressing the joint unobservability of investor AI adoption and managerial belief

[2] Disclosure policy can also be adjusted to cater to different investors' demands (Cao et al. 2023). Quasi-natural experiments cannot easily hold constant different investors' demands for information and rule out alternative explanations. In particular, such shocks seldom produce precisely defined variation in information demand, nor do they accommodate heterogeneous adoption patterns across investors; both of these things are important for isolating firms' strategic responses.

[3] We use "information demand" to denote the quantity as well as the informational content investors require. However, the realized supply of disclosure is endogenous to managers' beliefs about how investors will process and interpret the information, over and above considerations of the amount or content per se. Alesina et al. (2018) manipulate the belief of intergenerational mobility and make perceptions of mobility more pessimistic in their interventional survey.

formation in archival settings.[4] In particular, each firm receives the same four standardized questions on current financial performance, future growth, current industry conditions, and future industry trends. The sole variation is the presentation format. In the control group (T0), questions use conventional investor language. In the treatment group (T1), questions adopt a robotic linguistic style and a "[Query-00*x*]" prefix to mimic an automated agent. This is in the same spirit as a nudging effect (e.g., Tversky and Kahneman 1981; Chetty 2015; DellaVigna and Linos 2022). Where feasible, we further reinforce the chatbot identity with a robot profile image as an explicit AI label (T2). Essentially, this is a combination of nudging with labeling (T1+T2).[5]

One may be concerned that our treatment does not vary perceived AI identity as expected. We perform two manipulation checks to alleviate this concern. First, we run a lab experiment in which participants are significantly more likely to flag an AI-styled questioner as a robot, algorithm, or AI agent—an effect driven primarily by the robotic language and reinforced by the icon. Reassuringly, the manipulation does not lower perceived trustworthiness and, if anything, raises perceived intelligence and credibility, which alleviates the concern that our treatment merely signals an untrustworthy or differently sophisticated questioner rather than an AI-equipped one. Second, we employ a large language model (LLM) to verify that the treatments vary only in perceived identity and not in unintended dimensions of phrasing. We report the lab experiment alongside the LLM-based check in Online Appendices A and B.

The experiment ran from April 14 to June 6, 2025, covers 2,722 firms, and aligns with the annual report disclosure window. We tie our experiment to the annual report season to minimize

[4] By ensuring that both treatment and control managers respond to the same underlying news, our control group provides a credible behavioral baseline (List 2011).

[5] Generative AI can closely mimic human narrative, making it difficult for individuals to distinguish AI-generated text from human-authored content in the absence of explicit cues (Yin et al. 2024). Our objective is not to perfectly replicate AI output but to induce the belief that an AI bot is present. To that end, we include explicit prefixes and labels in addition to wording. Though these may deviate from how real GenAI operates, they resemble a chatbot and trigger the perception of AI involvement.

uncertainty about the arrival of private information (Dye 1985). Additionally, by concentrating sample conferences within a relatively short period, we effectively mitigate the concern of cross-subject contamination and ensure the stable unit treatment value assumption (SUTVA).

Because OEC presentation dates are revealed gradually, we implement contingent rolling randomization. Analogous to Chan et al. (2024), we rely on independent coin tosses without stratification: each day, we collect newly announced OECs from the regulator-designated *cninf* channel, form date-based batches, and, on the day before each firm's conference, randomly assign to the treatment or control group with equal probability. This procedure maintains unbiased allocation despite the unstratified nature.

We show that managers are 13.2% less likely to respond to AI-styled questions compared to parallel human-styled questions from the control group, suggesting that the perception of questions from AI bots reduce the information provided by managers. Disaggregating the two treatment layers—linguistic nudging and explicit labeling—we find that language nudging alone (T1) reduces response rates by 6.4%, while the addition of explicit AI labeling (T2) generates a further 16.6% decline. The stronger effect of labeling is consistent with an "AI image" providing a clear, salient signal of AI involvement. By contrast, linguistic nudges are sufficiently subtle to leave managers uncertain about AI readership, eliciting relatively modest disclosure adjustments. Furthermore, we find that the non-response effect is stronger for industry-related questions.

Next, we leverage the exact respondent identity and link disclosure choices to manager- and firm-level traits and uncover heterogeneity in responses to AI-styled questions. We find that the decline in response rates is attenuated for state-owned enterprises (SOEs), larger for female and older managers, and most pronounced when the manager has a computer science background. These patterns are consistent with non-market incentives of SOEs (Shleifer and Vishny 1994).

They are also consistent with the fact that female managers exercise greater caution with communication and senior ones are more concerned with perceived reputational risk (Hambrick and Mason 1984; Huang and Kisgen 2013; Serfling 2014). The patterns are also consistent with increased awareness of AI's superior capabilities among managers with a tech background.

As a rapidly evolving technology, AI can be interpreted as either a positive or negative economic force: it can be as a tool that enhances information processing (Sims 2003, 2006; Bertomeu et al. 2025a; Blankespoor et al. 2020; Kim and Kim 2025) and reduces the disclosure regime by making manipulation more costly, or it can be an "alien" presence that triggers distrust (e.g., Luo et al. 2019; Tong et al. 2021; Yin et al. 2024). Either channel, among others, may lead managers to reduce the information they provide. We deem our cross-sectional heterogeneity evidence as more consistent with the interpretation of AI's disciplinary role arising from enhanced information processing capabilities.

To better understand the underlying mechanisms, we focus on the more empirically tractable channel: AI's role in enhancing investors' information processing, thereby constraining managers' ability to create a flexible narrative. This choice is driven by empirical feasibility rather than theoretical priors.[6] To this end, we rely on the setting that enables us to track shifts in the managerial narrative, which is a key indicator of disclosure management (Healy and Palepu 2001; Beyer et al. 2010; Huang et al. 2014; Baik et al. 2025), and to exploit the real-time nature of investor questions and managerial responses using intraday trading data.

We measure the sentiment of managers' responses using FinBERT1, an academia-established language model trained on the Chinese financial corpus (Lu et al. 2025; Zhang et al. 2025). Following Jha et al. (2025), we construct sentiment axes by manually drafting positive and

[6] We note that other sources of AI aversion are difficult to isolate, and we cannot construct a human investor with AI-equivalent capabilities.

negative benchmark responses for each question, embedding them with FinBERT1, and averaging them into positive and negative "poles." Each response is then scored by its cosine similarity to this axis, allowing for a context-specific and theoretically grounded measure of sentiment that avoids the noise of generic lexicons. Consistent with reduced narrative management (Huang et al. 2014), managers exposed to perceived AI bots provide responses with significantly lower sentiment scores and a more neutral narrative relative to the control group.[7]

We also find that, conditional on responding, managers provide longer answers to AI-styled questions in a timelier manner. This intensive-margin pattern is difficult to reconcile with a pure "anti-AI" or distrust channel, which would predict disengagement and less-effortful communication (e.g., shorter and more delayed responses) among those who participate. Instead, it aligns with the disciplinary mechanism: perceived AI-enabled monitoring raises the expected cost of selectively framed narratives, inducing managers who do respond to exert greater effort.

Lastly, a distinctive feature of the livestream setting is that both investor questions and managerial responses are posted live on the platform and observable to market participants. From a market microstructure perspective, this allows us to isolate investor beliefs about the informational signal of responses, conditional on the perceived source (AI-equipped investors) of the question.[8] When AI-styled questions arrive, investors revise beliefs about their potential counterparties and about the expected informativeness of subsequent answers, and we indeed observe a decline in trading activity upon question arrival, which is consistent with investors either seeking to avoid trading against sophisticated AI-equipped counterparties or valuing the option to

[7] An alternative explanation is that managers use a negative narrative toward industry peers to make their own performance appear stronger. However, as discussed in Section 6.2, we find little support: peer-related sentiment is generally positive, and responses to AI-styled questions show increased, not decreased, neutrality.

[8] This aligns with the k-level thinking in which agents form beliefs based on iterative reasoning about others' beliefs. For example, a level-0 agent may act follow a heuristic; a level-1 agent best responds to level-0; a level-2 agent best responds to level-1, and so on. See Nagel (1995) and Camerer et al. (2004).

wait for more-informative managerial responses.[9] When the corresponding managerial responses are released, we find no differential reaction in abnormal volume, a pattern consistent with AI both reducing belief dispersion (which dampens trading) and raising disclosure thresholds (which strengthens signals), so that these two forces largely offset each other in the response window.

In sum, our RCT shows that perceived AI bots reduce managers' voluntary disclosure. While a decline in disclosure is sometimes viewed negatively (e.g., Boone and White 2014; Bird and Karolyi 2016), our evidence suggests, in this RCT context, a shift in the composition of information supplied: AI presence appears to suppress discretionary narrative management. In particular, when managers perceive that AI bots can more effectively debias disclosures, the effective disclosure threshold rises. As a result, managers become less likely to respond, reducing the overall supply of information, but in a positively selected manner, as only firms with sufficiently favorable information choose to disclose.[10]

2. Literature Review

The classical theoretical literature on voluntary disclosure, beginning with Grossman (1981), establishes the benchmark unraveling result: when all managers are informed of the true value of their firms, disclosures are costless and verifiable, and investors update beliefs rationally, even the low-type firms have an incentive to disclose to separate themselves from the worse firm. This leads to full unraveling, in which all firms disclose truthfully, and investors interpret silence as worst-case news. The critical assumptions underpinning this result are: (1) all managers observe their private type, (2) disclosure is truthful and frictionless, and (3) investors apply worst-case

[9] If investors anticipate that responses to AI-styled questions will be more informative, the option value of waiting also increases (McDonald and Siegel 1986; Acharya et al. 2011), potentially leading to a delay in trading activity until the response is released.

[10] While our analysis emphasizes the disciplining role of AI, we acknowledge that reduced disclosure may also reflect the "dark side" of AI, such as negative sentiment or distrust toward AI bots.

inferences upon observing silence.

However, the full unraveling result is fragile. The follow-up research demonstrates that small frictions can generate threshold (cutoff) disclosure equilibria (Verrecchia 1983; Dye 1985; Jung and Kwon 1988). For example, when the arrival of private information is uncertain or unobservable (i.e., not all managers are informed), managers with moderately bad news may prefer to pool with worse or uninformed types of firms rather than reveal and suffer the reputational consequences. Similarly, proprietary or reputational costs to disclosure can also lead to selective silence. In these settings, disclosure follows a cutoff rule: the manager discloses if and only if the private signal is above the disclosure threshold, and silence becomes strategic.

The recent emergence of AI introduces a new inference channel that has not been fully addressed in these classical models. AI-based tools can enhance information processing by accurately summarizing disclosed content, identifying anomalies, or inferring hidden cues (Sims 2003; Dong et al. 2016; Blankespoor et al. 2020; Kim and Kim 2025; Chang et al. 2025; Bradshaw et al. 2025). But they may also introduce hallucinations, complicating how managers perceive the payoff from disclosure (e.g., Bertomeu et al. 2025b). Therefore, the introduction of AI has the potential to fundamentally shift the tradeoff faced by managers, changing both the disclosure threshold and the extent of manipulation.

In parallel, a large empirical literature focuses on the demand-side forces behind disclosure, especially the role of institutional investors. Studies show that firms with greater institutional ownership and attention are associated with more timely, transparent, and richer disclosures (e.g. Healy and Palepu 2001; Beyer et al. 2010; Boone and White 2015; Bird and Karolyi 2016; Abramova et al. 2020). These findings suggest that institutional investors demand more disclosure and enforce it through engagement, voting, and threat of exit. This empirical strand abstracts from

the manager's strategic disclosure game and treats disclosure as a simple response to quantity of information demand. Since institutional investors are also early adopters of AI tools, their presence may mask the distinct effect of AI inference mechanisms on strategic disclosure behavior.

To isolate the unique role of AI-mediated inference in shaping managers' disclosure choices, we design an RCT that carefully controls for multiple confounding forces. First, we equalize investor demand across treatment and control groups, ensuring that disclosure decisions are not confounded by heterogeneous attention or monitoring. Second, we implement the RCT in the context of earnings conferences tied to mandatory annual reports, where information arrival is certain and not subject to the manager's discretion in timing (Dye 1985). Third, we design our disclosure prompts to avoid any correlation with proprietary or sensitive information, ruling out the proprietary cost-based mechanism (Verrecchia 1983). By removing these confounders, we identify the effect of the manager's perception of investors' AI use on their willingness to disclose information and on their narrative decisions—a disclosure dimension that fits our experimental setting (Huang et al. 2014). This approach builds on recent theoretical work that highlights the role of AI in shaping managers' disclosure incentives (e.g., Bertomeu et al. 2025b; Kim and Kim 2025).

Our paper contributes to several streams of literature. The first stream studies AI as an information-processing technology in capital markets, documenting gains in decision-making by retail investors, professional analysts, hedge funds, etc. (Chang et al. 2025; Bertomeu et al. 2025a, 2025b; Cheng et al. 2025; Sheng et al. 2025; Cao et al. 2023).[11] However, how managers form beliefs when their audience includes AI bots remains underexplored. We contribute by providing causal evidence that perceived AI presence shifts the *supply* of voluntary disclosure, extending the disciplining role of AI beyond mandatory formats, such as 10-K filings (Cao et al. 2023).

[11] On the contrary, Croom (2025) shows that the interactivity of generative AI strengthens investors' self-assessments, highlighting one potential cost of using AI tools.

A second stream of literature uses microstructure settings to examine disclosure at high frequency, showing that real-time, text-based interactions reveal how markets process managerial communication intraday (Mayew et al. 2020; Rennekamp et al. 2022; Baik et al. 2025). Missing is causal evidence on how *AI-mediated* interactions simultaneously shape managerial narratives and immediate investor reactions. We contribute by showing suggestive evidence of more-neutral managerial narratives and delayed investor responses, which are consistent with larger belief-updating by managers and investors.

A third stream of literature exploits China's institutional settings for field experiments on disclosure, including credit registry changes, peer information avoidance, interactive investor platforms that raise demand for specific information, and AI-generated summaries that alter the informational environment (Liao et al. 2023; Chan et al. 2024; Wong et al. 2024; He and Xiahou 2025; Wong et al. 2025). Interventions in this literature primarily stimulate specific demand or change information format, leaving open how beliefs about an AI audience affect disclosure when demand and content are held constant. We contribute by isolating belief formation as the mechanism, showing how perceived AI usage affects disclosure supply and market reactions in a live, regulated conference setting.

Finally, the nudging literature shows that framing shifts beliefs and behavior without altering the choice set, offering a blueprint for belief-focused identification (Tversky and Kahneman 1981; Chetty 2015; Alesina et al. 2018). Disclosure studies, however, often face confounders from heterogeneous attention, timing discretion, and proprietary-cost motives (Verrecchia 1983; Dye 1985). We contribute to the literature by adopting a belief-focused design that addresses these confounders and show that investor adoption of AI discourages managerial communication via positive sorting, whereby only firms with sufficiently positive news choose to engage.

3. Experiment Protocol

3.1 Institutional background

In China, firm–investor communication takes place through a variety of channels. Formal interactions, such as earnings calls and roadshows, are typically restricted to invited professional participants. In parallel, more informal and less publicly visible meetings occur, including site visits organized by analysts or institutional investors. These traditional channels, however, are characterized by limited investor participation and selective access. In recent years, regulators have actively promoted online interactive investor platforms, which provide retail investors with direct, on-demand access to public firms. Through these platforms, investors can submit questions at any time, significantly broadening investor participation and enhancing the transparency of firm communication.[12]

Our study focuses on one specific form of online interaction: *online earnings conferences*. First introduced voluntarily around 2002, OECs have become a prominent feature of the investor communication landscape in China.[13] In line with the regulatory push for accessible and inclusive engagement between firms and investors, the China Securities Regulatory Commission (CSRC) issued guidelines in 2022 encouraging listed firms to conduct earnings conferences through various formats, including online text-based chats.[14] Firms retain discretion over the choice of

[12] Although channels like earnings conferences, site visits, and investor meetings provide opportunities for enhanced information access, they are typically limited to invited institutional participants. Even when retail investors are permitted access, they often lack the necessary resources or incur higher costs to participate effectively. This disparity is particularly salient in China, where retail investors account for approximately 85% of daily trading volume (Jones et al. 2021), suggesting that a large portion of trades occur with disadvantaged informational positions. Chinese regulators have consistently aimed to bridge the informational gap between retail and institutional investors. The China Securities Regulatory Commission has explicitly required equitable information disclosure during earnings conferences, roadshows, and site visits since 2007. Further addressing the issue, the Shenzhen Stock Exchange and the Shanghai Stock Exchange introduced two interactive online investor platforms, EasyIR and eIR, in 2010 and 2013, respectively. These two platforms have gained wide popularity among investors, and their positive impacts on corporate governance and the overall information environment are well documented in the literature (Blankespoor 2022; Lee and Zhong 2022; Wong et al. 2024; He and Xiahou 2025).

[13] See: https://ir.p5w.net/roadshow/.

[14] See: http://www.csrc.gov.cn/csrc/c101954/c2334702/content.shtml.

hosting platforms, including commercial services such as ValueOnline, as well as official exchange-affiliated platforms like EasyIR's Cloud Interview and SSE Roadshow.[15] Additionally, the 2022 CSRC guidelines require listed firms to host an earnings conference after releasing their annual report; thus, the annual report earnings conference has become a time when firms face mandatory disclosure.

There are several unique features of OECs in China. First, in contrast to traditional audio calls used internationally, OECs are typically hosted as a livestream, along with online textual chats.[16,17] Second, in these OECs, both institutional and retail investors, once signed up, can freely submit real-time questions during the short conference window. Third, regulators explicitly promote OECs as a means to enhance transparency and facilitate direct engagement between investors and firm management, with an emphasis on information disclosure and investor education. As a result, the default expectation is for firms to respond promptly and substantively to all questions, including those that are basic or repetitive. However, firms still retain great discretion over the content and wording of responses (e.g., firms can give evasive answers or manipulate their narratives), making this setting more of a voluntary disclosure regime.

Importantly, the design of OEC platforms allows for customization of both linguistic style (nudging) and profile identity (labeling). We exploit these flexibilities to experimentally simulate

[15] Please refer to the website for the corresponding platform: ValueOnline (https://www.valueonline.cn/), the Cloud Interview board (https://irm.cninfo.com.cn/views/interview) of EasyIR and SSE Roadshow (https://roadshow.sseinfo.com/home).

[16] Our setting is quite different from the traditional earnings conference call setting. China's legal system has a concept called *analyst meeting*—conferences held for financial analysts—which is what are typically called "earnings conferences" internationally. For example, the current Regulation on Disclosures of Listed Firms by CSRS states, "35. Listed firms should not provide insider information to any entity or any person when holding earnings conferences, analyst meetings, roadshows, site visits or any other forms of communication." (http://www.csrc.gov.cn/csrc/c101953/c7547359/content.shtml). Analyst meetings are closer to earnings conference calls in the United States as they are typically targeted to analysts and held via audio call. But there is one key difference between analyst meetings in China and earnings conference calls in the United States: at least from the regulators' perspective, analyst meetings are not necessarily related to and held shortly after earnings announcements.

[17] Although regulators impose no requirements on the format of earnings conferences, nearly all earnings conferences in our sample contain online textual chats (see Table 1).

AI-like investor inquiries—questions that are written by us but styled to resemble those generated by automatic tools (e.g., rule-based chatbots or LLMs like DeepSeek). By altering both the linguistic style and presentation (i.e., question phrasing and profile visuals), we nudge managerial perceptions about the identities of the questioners. Such fine-grained manipulations of perceived investor identity are difficult to implement in archival or quasi-natural settings.

In addition, the textual and publicly accessible nature of OECs provides rich documentation, enabling precise, real-time tracking and coding of both investor questions and managerial responses. This structure allows us to isolate the value relevance of highly specific pieces of information disclosed by managers in direct response to questions posed by our two types of experimental questioners. Because these disclosures occur in real time, they can be immediately incorporated into market activities, facilitating a clean empirical mapping between the nature of investor query, the corresponding managerial disclosure, and the market reaction. While alternative communication channels, such as private meetings or earnings calls, may also generate managerial disclosures, they typically require privileged access and are not uniformly observable. By contrast, the public, standardized, and scalable structure of OECs provides a transparent setting that is particularly well-suited for large-scale experimental implementation at relatively low cost.

3.2 Randomization and sample construction

Our sample consists of OECs held between April 14 and June 6, 2025, spanning eight weeks and covering 37 distinct calendar days.[18] Our sample period covers around 73% of earnings conferences for the 2024 fiscal year. As Figure 1 shows, the dates of earnings conferences for the 2024 fiscal year within our sample period follow a normal-like distribution, with a majority of earnings conferences held in the middle of our sample period; May 9 is the peak day with over 250

[18] We conducted the experiment after obtaining the IRB approval and therefore missed earnings conferences held prior to April 14.

earnings conferences.[19]

We conduct randomization via independent coin-tosses, a mathematic structure similar to the group assignment process in Chan et al. (2024) that applies when a subject pool cannot be clearly predefined.[20] In our setting, the schedule of the OECs is not predefined since certain details, such as when an earnings conference is held (i.e., the exact date and duration), whether it is online or offline, and which platform is used for online conference, are largely at the firm's discretion, and firms usually do not disclose their scheduled presentations upfront. Further, earnings conferences are distributed unevenly across dates, and in some extreme cases, there can be more than 250 conferences held on one single day (see Figure 1).

We choose the coin-toss design and arrange the toss game as follows. For each day, we collect all the newly released announcements of earnings conferences from the cninf website (https://www.cninfo.com.cn/), a regulator-designated online channel of filings and disclosures for all listed firms in China. We check these announcements and keep basic information including conference dates. Thus, each conference is naturally blocked into one of the date batches.

On the day before the conference date of each batch, we independently toss a coin for each of the conferences in the batch.[21] By design, each conference independently receives one of the two treatments with an identical probability of 0.5.

Table 1 outlines the sample screening procedure. We start with the 3,481 earnings conferences announced to be held between April 14 and June 6, 2025, involving a total of 4,049 unique firms.[22]

[19] Zero values in this figure reflect the fact that firms do not hold earnings conferences on weekends or national holidays. Note that April 27, 2025, was a Sunday but was treated nationally as a weekday to compensate for five-days off starting from the International Labor Day on May 1. Only a few earnings conferences are held on April 27.

[20] In Chan et al. (2024), each subject, or data point, is an event in which a person receives a message from a charity campaign. The display format of this message was randomized.

[21] We forgo conferences announced on the days they are held. To (strategically) announce an earnings conference very shortly before it is held is an unwelcome practice from the perspective of regulators because it takes time for announced information to be fully disseminated to and digested by retail investors.

[22] In some cases, one conference can be jointly held by multiple firms associated with the same business scope or

First, we exclude 33 non-OECs that are held offline or via phone call for which we cannot exert our treatments. Second, for each firm, we retain only its earliest OEC. This is to avoid within-firm contamination and ensure SUTVA in cases in which a firm holds multiple OECs during our sample period; we exclude 52 OECs. Third, we exclude 58 joint earnings conferences (JECs) in which multiple firms respond to investor inquiries at the same conference. Fourth, we restrict our sample to OECs held on one of the nine selected platforms that are suitable for RCT implementation (see Table 2 for more details), on which firms generally do not verify the identities of OEC attendees, and on which attendees generally attend OECs without firms' approval. We exclude 67 OECs held on other minor platforms that have specific mechanisms to screen the credentials of participants, which may violate our identification assumption.[23] Fifth, we exclude five OECs that involve an offline venue where the management focuses solely on in-person interactions and does not engage with any online textual questions. Sixth, we restrict our sample to OECs held solely for annual reports or for both annual and Q1 reports; this reduces our sample by 108 OECs.[24] Seventh, we exclude 436 OECs that were unable to be part of the experiment due to institutional frictions, such as reporting errors in the disclosed conference schedule or functional problems in their websites.[25] This procedure yields a final sample of 2,722 OECs hosted by 2,722 unique firms.

3.3 The assignment of treatments

We seek to examine whether and how corporate managers respond differently to questions

---

the same enterprise group (typically state-owned entities). We delete those collective conferences in the third step of sample selection to prevent contamination.

[23] Some minor platforms such as Comein Finance (https://www.comein.cn/home/index) are targeting a special set of investors. Typical hosts of OECs on these platforms screen the credentials of participants ex ante and block access of participants that they suspect are not authentic.

[24] Due to overlapping disclosure schedules, OECs sometimes simultaneously address annual and Q1 reports. Although not mandatory, firms sometimes hold separate conferences dedicated exclusively to quarterly or semi-annual reports; however, for clarity, we exclude these from our sample.

[25] Reasons for such failures include firms' erroneous announcements of their OEC date and announcements on the very days of a presentation.

that appear to come from AI, as opposed to questions that resemble those posed by human investors, during OECs. Our experimental protocol comprises a control group and a treatment group, with randomized assignment across OECs conducted on nine leading investor Q&A platforms. These platforms differ greatly in a variety of dimensions; we outline these differences in Table 2.

For each control group OEC (T0), we submit four questions that resemble those commonly asked by retail investors. These questions address four core themes: (1) the firm's current financial performance, (2) the firm's long run growth potential, (3) current industry performance, and (4) long-run industry outlook. Each question targets one thematic dimension and is phrased in a natural, conversational way to reflect average participant behavior. Panel A of Figure 2 provides an example of a control-group question from SSE Roadshow.

In the treatment group (T1), we submit questions that address the same four themes and follow the same sequence. However, the treatment questions are intentionally phrased in a robotic and formalized style, designed to resemble the linguistic patterns of AI-generated queries (e.g., rule-based chatbots or LLMs). This linguistic shift serves as a nudge: it preserves the informational content of the question while subtly signalling that the source may be non-human. In addition, each treatment question includes a prefix "[Query-00*x*]", where *x* denotes the question number. This prefix reinforces the impression of a machine-generated question. Panel B of Figure 2 illustrates this treatment condition. Note that the purpose of the treatment here is not to fully mimic generative AI but to insert a perception that the questioners are AI-equipped.

To further strengthen the salience of AI identity, we implement an additional labeling mechanism when permitted by the platform interface (T2). Specifically, on platforms that allow users to modify their profile images (such as Quanjing), we upload a chatbot avatar before submitting the questions. Panel C of Figure 2 provides an example of this visual cue. This enables

us to vary treatment intensity across OECs: the baseline treatment consists of the AI nudge (T1), while the augmented treatment combines both textual and visual labeling of a chatbot identity (T1+T2).

We outline our experiment design in Figure 3. Additionally, Table 3 provides the detailed wording of the experimental questions and the profile image for the T2 treatment.[26]

3.4 Validation of treatments

Validating the manipulation is important because managerial perceptions are not directly observable in the field and cannot be checked within the RCT itself. We therefore conduct two complementary exercises, reported in Online Appendices A and B, which we interpret as suggestive rather than definitive.

In the first exercise (See Online Appendix A), we conduct an IRB-approved online lab experiment[27] with 240 participants recruited from Credamo. Participants view a mock investor discussion thread modeled on Guba, a leading Chinese retail-investor platform, in a 2×2 between-subjects design that independently varies linguistic style and profile image, allowing us to isolate the robotic icon (T2) from the robotic language (T1)—a separation infeasible in the field. Both layers significantly raise the probability that participants flag the manipulated user as a robot, algorithm, or AI agent, with the language as the primary driver and the icon a smaller incremental effect. Importantly, the manipulation does not reduce perceived trustworthiness, and linguistic style alone does not shift perceptions of intelligence or credibility, mitigating the concerns that our

[26] In an ideal setting, we would impose Treatments T0, T1, and T1+T2 within the same OEC and on the same group of managers, allowing us to identify within-subject variation. However, this is not practically feasible due to two key constraints: (1) the duration of each OEC session is limited to 30–120 minutes, and (2) the average number of questions raised per session is around 11. (See Table 4.) Introducing multiple treatments within a single session risks not only crowding out other participants' questions but also generating interference among our own treatment arms. Therefore, we rely on randomization across sessions to ensure that, with a sufficiently large sample, between-OEC comparisons yield inferences equivalent to that from within-OEC treatment variation.

[27] This lab experiment was independently reviewed and approved by the Institutional Review Board of Renmin Business School, Renmin University of China (2026R24).

treatment merely signals an untrustworthy or more sophisticated investor. Because we cannot fully replicate an OEC or recruit listed-firm managers, we read this evidence as suggestive, under the assumption that managers do not respond systematically differently from the general population.

In the second exercise (See Online Appendix B), we ask a state-of-the-art generative LLM, DeepSeek-R1, to rate, for each of the four thematic question pairs, the likelihood that the human-styled (T0) and AI-styled (T1) versions were machine-generated. The model consistently classifies the human-styled questions as likely human-authored and the AI-styled questions as more likely machine-generated, attributing the latter to the "[Query-00x]" prefix, the formal salutation, and the templated phrasing.

4. Data

4.1 List of datasets

Our first dataset is derived from our experimental implementation and consists of 2,722 OECs, each conducted by a unique firm. For each OEC, we submit four experimental questions, yielding a total of 10,888 firm (conference)-question observations. The second dataset comprises firm-level characteristics obtained from the Chinese Stock Market and Accounting Research database. Our third dataset contains intraday trading data for all firms listed on the Shenzhen and Shanghai Stock Exchanges during our sample period, from April 14 to June 6, 2025.[28] These data are from the SinoFin CCER database.

4.2 Measuring sentiment

To construct our sentiment measure, we use FinBERT1, a domain-specific language model tailored for Chinese financial text.[29] Introduced in 2020 by a FinTech firm, ValueSimplex,

[28] Our OEC sample does include firms listed on the Beijing Stock Exchange (BJSE). But intraday trading data for these BJSE-listed firms are unavailable.
[29] See: https://github.com/valuesimplex/FinBERT/blob/main/FinBERT1_README.md.

FinBERT1 is pretrained on a large corpus of Chinese financial documents, including news articles, analyst reports, corporate filings, and Wikipedia entries. The model adopts the same underlying architecture as Google's original BERT framework, and has demonstrated strong performance in understanding financial text (e.g., Lu et al. 2025; Zhang et al. 2025).

We measure the sentiment of management responses following the approach proposed by Jha et al. (2025). For each of the four experimental questions, we manually construct a set of benchmark responses that directly address the question content, with some expressing clearly positive sentiment and others clearly negative sentiment. For example, in response to the question regarding recent financial performance, representative positive and negative responses might include "We outperformed expectations" and "We underperformed expectations," respectively. (Please refer to Appendix C for details of predefined benchmark responses.) These manually written sentences are designed to be direct and non-evasive. Each sentence is embedded into a 768-dimensional vector space using FinBERT1. For each question, we compute the positive pole by averaging the embeddings of positive benchmark sentences, and we compute the negative pole by averaging those of negative benchmark sentences. This procedure yields a question-specific sentiment axis, defined as the vector difference between the positive and negative poles. Formally, we define the sentiment axis as:

$$Axis_q = \text{avg}\left(\left\{\text{emb}(s): s \in \mathcal{P}_q^+\right\}\right) - \text{avg}\left(\left\{\text{emb}(s): s \in \mathcal{P}_q^-\right\}\right), \quad (1)$$

where q denotes one of our experimental questions; $\text{emb}(\cdot)$ is a FinBERT1 transformation that embeds a text string (i.e., a sentence) s to a 768-dimension vector; $\mathcal{P}_q^+$ and $\mathcal{P}_q^-$ are the predefined sets of positive and negative benchmark responses to question q, respectively; and $\text{avg}(\cdot)$ is a function to compute the equal-weighted average vector of a vector set.[30]

[30] A model of BERT architecture can typically receive either one or multiple strings at one time, but for ease of

With a predefined sentiment axis for each of our four questions at hand, we rely on a cosine similarity metric to obtain the final sentiment scores for management responses. We first parse each response to question q into sentences based on the occurrences of periods, exclamation points, question marks, and colons—a rule-based method. Next, for each sentence s of a response to question q, we derive its FinBERT1 embedding and calculate its cosine similarity with corresponding sentiment axis $Axis_q$,

$$\cos\big(\text{emb}(s), Axis_q\big) = \frac{\text{emb}(s) \cdot Axis_q}{|\text{emb}(s)| \times |Axis_q|}. \tag{2}$$

The sentiment score for a response to question q on conference c is then

$$Sentiment_{q,c} = \frac{1}{N_{q,c}} \sum_s \cos\big(\text{emb}(s), Axis_q\big), \tag{3}$$

where $N_{q,c}$ is the count of identified sentences of responses to question q in OEC c. Sentiment is bound between –1 and 1. A larger value indicates a more positive sentiment.

Note that we retain the full embedding vectors produced by FinBERT1, allowing us to capture finer-grained variations in sentiment. In our setting, management's online responses are often brief, consisting of only a few sentences. In such cases, traditional proportion-based sentiment measures constructed at the paragraph or article level can be uninformative (e.g., Cao et al. 2023; Huang et al. 2023; Rajan et al. 2023). In addition, this measure captures an additional dimension of managerial narrative: sentiment neutrality. While cosine similarity values close to 1 or –1 indicate strong alignment with the positive or negative poles, respectively, values near zero, at least in part, reflect sentiment that is close to neutral. Thus, this framework allows us to quantify not only the

illustration, we present our procedure here as if it receives only one each time. The model first parses the received string (i.e., a sentence) into several tokens via a built-in tokenizer and then returns a list of embeddings corresponding to each of all these tokens, along with a [CLS] embedding that represents the whole received string. We are interested only in this [CLS] embedding. It should be noted that the received string does not need to have appeared in its pretraining corpus, which makes it work well out of the box and out of sample.

direction but also the degree of affective content, or its absence, within very short textual responses. We define our measure of sentiment neutrality as follows:

$$Neutrality_{q,c} = -\frac{1}{N_{q,c}} \sum_{s} \left| \cos\left(\text{emb}(s), Axis_q\right) \right|. \quad (4)$$

We add a negative sign so that a greater value indicates more-neutral sentiment. Although a close-to-0 value of cosine is regarded as a proxy for sentiment neutrality.

4.3 Control variables

To alleviate the concerns that our results are driven by firm characteristics that potentially covary with our unstratified treatment assignments, we follow Abramova et al. (2020) and control for a set of variables that can impact the propensity of managers to voluntarily disclose information. These controls are market capitalization (Size), book-to-market ratio (Book to Market), financial leverage (Leverage), annual stock return (Annual Return), volatility of monthly stock returns (Return Volatility), whether or not the firm is at a loss (Loss), whether or not the firm experiences a rise in annual earnings per share (EPS Rise), absolute value of relative change in annual earnings per share (Abs. ΔEPS), number of analysts issuing earnings estimates (Analysts Following), and institutional ownership (Inst. Ownership). We also control for a set of transparency indicators from He and Xiahou (2025), including proportion of independent directors in all board members (Board Independence), ratio of R&D expenditures to revenues (R&D Intensity), whether or not the firm is owned by the state (SOE), sales to the top five customers (Customer Concentration), and purchases from the top five suppliers (Supplier Concentration).

Given the wide latitude that firms have in framing corporate disclosures during OECs, their default strategy is to emphasize good news and withhold bad news. Accordingly, management responses are generally expected to exhibit positive sentiment, and the presence of negative sentiment can signal reduced message management. To isolate this strategic dimension, in

sentiment analysis of managerial responses to our questions, we control for a set of firm-level sentiment predictors established in the literature (Huang et al. 2014), as variation in sentiment may also reflect underlying firm fundamentals rather than manipulation.[31] This allows us to absorb the fundamental component of sentiment and focus on the abnormal, potentially strategic, variation, closer to what we define as signal manipulation.

5. Results

5.1 Summary statistics and balance checks

Table 4 reports the descriptive statistics for our sample. Panel A summarizes firm-level characteristics across the 2,722 OECs. The firms included in our sample account for about half of the total number of firms in the Chinese stock market, amounting to a total market cap of 32 trillion yuan. On average, firms in our sample are about 24 years old, with a mean market capitalization of 12 billion yuan and a book-to-market ratio of 0.55.[32] Profitability is modest, with a mean ROA of 1.4%, and annual stock returns are slightly negative at –1.4% on average. Firms exhibit a return volatility of 14.9% and an earnings surprise close to zero, and around 25% of firms are SOEs. In terms of governance and operations, independent directors account for 38.4% of all board members, and sales to the top five customers (Customer Concentration) and purchases from the top five suppliers (Supplier Concentration) account for 37% and 35% of the total sales and purchases, respectively. Additionally, our sample OECs contain an average of 11 real-world textual Q&As, and last for around 1.5 hours.[33] On average, our sample firms respond to three of our experimental

[31] We control for firm age (Age), market capitalization (Size), book-to-market ratio (Book to Market), annual stock return (Annual Return), volatility of monthly stock returns (Return Volatility), return on assets (ROA), volatility of return on assets (Performance Volatility), whether or not the firm is at a loss (Loss), annual change in return on assets (ΔROA), earnings surprise (Earnings Surprise), and consensus earnings per share forecast by analysts (Analyst Forecast). We do not include BUSSEG and GEOSEG as in Huang et al. (2014) for lack of necessary data in China.

[32] As market capitalization distribution is highly right-skewed, we take the natural logarithm to compute Size in our analysis sample. Nevertheless, we report the original value mean here in that re-exponentiation of the mean of the natural logarithm can yield a largely inconsistent figure.

[33] Two sample OECs do not have an explicit end time. Moreover, we do not have access to the original webpages of

questions, and around 12% of firms do not answer any of our questions. Panel B of Table 4 focuses on the firm-question level, showing that about 74% of experimental questions receive managerial responses. Response sentiment is positive on average ($mean = 0.08$), while the neutrality measure suggests that many answers are strategically affective. The average response takes around 35 minutes, contains about 183 Chinese characters, and spans roughly four sentences.[34]

Across the set of pre-experiment firm-level controls, the treatment and control groups appear highly comparable. Table 5 presents balance checks between the two groups using Weltch's *t*-test. We adopt the procedure introduced in Benjamini and Yekutieli (2001), which downward-adjusts critical values to identify statistical significance, to control for the false discovery rate (FDR) in multiple comparisons. All differences in means are statistically insignificant, consistent with successful randomization. Overall, the randomization procedure yields well-balanced experimental groups, mitigating concerns that the subsequent results are driven by pre-existing systematic differences in firm characteristics. Nevertheless, we explicitly control for all these characteristics in our formal tests.

5.2 Managers' likelihood of responding

We begin with univariate comparisons. As a benchmark, in Panel A of Figure 4, we first plot the response rates of our experimental questions. We observe an overall response rate of 74.4%, and managers are most reluctant to respond to Q3—it has a response rate of only 64.5%. Panel B of Figure 4 shows that treated questions, those written in AI-like style, are associated with a 15.1 percentage point decline in response rate relative to the control group benchmarks. This reduction

two sample OECs, resulting in missing values in # Q&As (the number of questions and answers excluding those we posed).

[34] We do not have documented timestamps of some of our sample firm-questions, resulting in missing values in Response Time. These primarily comprise firm-questions on Shanghai Securities News, where the timestamp of online questions and answers are not displayed on the webpage.

is evident across all four thematic categories, ranging from –11 to –17 percentage points, with the larger declines observed for industry-related inquiries. These differences are all statistically significant, providing initial evidence that perceived AI bots discourage managerial disclosure.

We next turn to Probit regressions to formally quantify the treatment effects:

$$Response_{q,c} = \alpha_j + \alpha_b + \alpha_p + \alpha_q + \gamma \times Treated_c + \theta \times X_c + \varepsilon_{q,c}; \quad (5)$$

$$Response_{q,c} = \alpha_j + \alpha_b + \alpha_p + \alpha_q + \gamma \times Text_c + \delta \times Text_c \times Image_c + \theta \times X_c + \varepsilon_{q,c}, \quad (6)$$

where q denotes question, c denotes firm (conference), j denotes industry, b denotes hosting date batch, and p denotes platform. α_j, α_b, α_p, and α_q are industry, batch, platform, and question fixed effects, respectively; $Response_{q,c}$ equals 1 if question q at conference c receives a reply; $Treated_c$ and $Text_c$ equal 1 for treated OECs; and $Image_c$ equals 1 for conference c in which we explicitly labeled as AI. X_c is a set of firm-level controls, including Size, Book to Market, Leverage, Annual Return, Return Volatility, Loss, EPS Rise, Abs. ΔEPS, Analysts Following, Inst. Ownership, Board Independence, R&D Intensity, SOE, Customer Concentration and Supplier Concentration. (See Table Variable Definitions for details.) When estimation is conducted in full sample, standard errors are clustered by firm, and when estimation is conducted respectively for each thematic category of question, by-firm clustering in fact yields robust standard errors because each firm (conference) receives each question only once.

Table 6 reports Probit estimates of the likelihood of response. The analysis sample in this table is all firm-questions with non-missing control variables. Across all four types of questions, managers are about 13 percentage points less likely to respond to treated questions. Relative to the sample average of 74 percentage points (Table 2, Panel B), this represents an 18% reduction in the response rate. The decomposition shows that AI-style nudging alone (T1) reduces responses by about 6 percentage points, while explicit AI labeling (T2) produces a much sharper decline of 17

percentage points. These magnitudes confirm that labeling has a stronger impact than subtle linguistic cues, suggesting that clear and salient signals of AI bots are particularly effective in shifting managerial behavior.[35] Panel B further shows that this negative treatment effect is remarkably consistent across all four questions, with estimates ranging from –9 to –15 percentage points. This suggests that the effect of perceived AI bots is broad-based rather than concentrated on specific types of information. In addition, the reduction in response propensity is more pronounced for industry-related questions. Our lab experiment supports this interpretation: in a 2×2 design that separately varies linguistic style and the chatbot icon (Online Appendix A), both layers raise the perceived likelihood that the questioner is an AI agent, with the icon adding incrementally on top of the language. This corroborates that the larger T2 effect reflects the added salience of explicit AI labeling rather than any change in the underlying information requested.

Finally, we test whether these results are confined to a particular platform or question. Figure 5 illustrates treatment effects by platform and question. The negative impact of perceived AI bots is observed across nearly all of the nine OEC platforms and four questions, and many estimates are statistically significant. The consistency across both platforms and questions underscores that the reduction in response likelihood reflects a general response to perceived AI presence rather than an artifact of any single institutional channel. We include platform fixed effects throughout, which absorb level differences in response norms across platforms. We note, however, that the explicit labeling treatment (T2) is feasible only on platforms that permit custom profile images (e.g., Quanjing); the labeling effect should therefore be interpreted with this platform-specific feasibility in mind.

[35] One might worry that our results reflect AI drafting tools (e.g., AI-assisted respondents replying less to AI-styled questions) rather than managers' disclosure choices. This explanation is not obvious: drafting assistance would more naturally lower effort costs and facilitate replying, and the within-response patterns we document (more-neutral narratives and longer replies) do not align with simple automated screening and withholding.

5.3 Robustness checks

A concern with our design is that managers may gradually learn about the experimental treatments by observing earlier OECs, leading to spillover effects across firms over time. Such learning would threaten identification by violating the SUTVA. We directly test for this concern.

Figure 6 plots the daily impacts of perceived AI bots from the start of the experiment on two outcomes: Response (Panel A), and Response Time (Panel B). Each point represents the estimated treatment effect for a given conference date batch, along with 95% confidence intervals. If spillovers were present, we would expect systematic trends in these coefficients, such as the effects becoming stronger or weaker as the experiment progresses. Instead, the estimates fluctuate around stable levels with no consistent upward or downward pattern across the 37-day window. These results mitigate the concern that managers adjust behavior over time in response to earlier OECs.

Another concern is that because the questions we raise are generic, there is a possibility that other participants raise similar questions and crowd out ours. In the experiment, we try to ask our experimental questions as soon as possible after each OEC begins. Nevertheless, we conduct a robustness check and find that our observed treatment effects do not intervene with the crowding-out effects from semantically similar questions. (See Table OD1 in Online Appendix D.) A related caveat is that some platforms allow firms to screen submitted questions before display, and we do not observe the full distribution of similar questions that other participants may have submitted but that were screened out. To the extent that managers condition their responses on the perceived popularity of comparable questions, this unobserved distribution is a limitation, although the crowding-out check above mitigates the most direct version of this concern.

Finally, the observed negative impact of perceived AI bots on response likelihood might also be driven by how intensely OEC attendees participate. The rationale is direct: intensive attendee

participation and question arrivals render any non-responses less salient to the audience, giving managers wider leeway to forgo questions they dislike, including our AI-styled ones. We deem this possibility highly unlikely in that our coin-toss process is independent of the ex-ante popularity of any OEC event that largely determines the intensity of the arrival of the questions. As is evident in Table 5, our randomization procedure yields comparable treatment and control groups. Nevertheless, we conduct a robustness check to explicitly rule out this possibility. (See Table OD2 in Online Appendix D.)

5.4 Cross-sectional heterogeneity in treatment effects

A distinctive advantage of our setting is that we observe the precise identity of the respondent. This feature allows us to link responses to both personal attributes of the responding manager and firm characteristics. To assess who is most sensitive to AI-styled questions, we interact the treatment indicator with four pre-specified cross-sectional markers: SOE ($=1$ for state-owned enterprises), Female ($=1$ if the respondent is female), Elderly ($=1$ if the respondent's age is in the top quartile of the sample), and Computer Science ($=1$ if the respondent has a computer science background).[36] This design tests whether respondent and firm traits amplify or attenuate strategic non-response to perceived AI readership.

Results in Table 7 show a consistently negative baseline treatment effect, indicating that AI-styled questions reduce response likelihood. The interaction terms reveal substantial heterogeneity. For SOEs, the interaction increases responsiveness by 7.2% relative to a baseline decline of 15.0%, yielding a net reduction of about 7.8%. For female managers, the interaction lowers responsiveness

[36] When the interacted marker is personal characteristics of the responding manager (i.e., Female, Elderly, or Computer Science), our analysis sample is restricted to OECs hosted on SSE Roadshow and Shanghai Securities News, the two Shanghai-based platforms, where we are required to explicitly select one respondent from attending managers before posting our question. On platforms other than these two, such selection is not a compulsory. Due to data availability issues, we have many missing values in Computer Science.

by 13.2% on top of a baseline drop of 5.1%, producing a net decline near 18.3%. Older managers experience a larger decrease, with the interaction reducing responsiveness by 10.5% and the baseline accounting for another 3.6%, for an overall decline of roughly 14.1%. The largest effect occurs when the CEO has a computer science background: the interaction reduces responsiveness by 46.9% in a subsample in which the baseline is close to zero, implying an overall decline of about 50%. Indicator estimates alone indicate that, absent treatment, females and those with a computer science background respond more frequently, but their responsiveness falls sharply when questions appear to be AI-styled.

Taken together, these patterns reveal interesting heterogeneity in managers' responses to perceived AI readership, plausibly related to pre-existing governance and managerial traits. The attenuated effect at SOEs is consistent with weaker capital market incentives to engage in discretionary narrative positioning (Shleifer and Vishny 1994). Larger declines among female and older managers align with greater caution in communication or perceived reputational risk (Hambrick and Mason 1984; Huang and Kisgen 2013; Serfling 2014). The strongest response among managers with a computer science background is consistent with those manager's having a greater awareness of AI's evaluative capabilities, raising the perceived cost of imprecision. We view all these interpretations as suggestive. We caution that we cannot verify that the executive we identify as the respondent personally drafts the answer; during an OEC, a team of executives and other employees typically assists in preparing responses. Nevertheless, because a specific executive is asked to answer and is publicly named, that executive bears responsibility for any errors or misinformation in the response. To the extent that the executive cares about his or her reputation and seeks to limit legal risk, he or she should at least be involved in drafting or monitoring the responses. We therefore treat the respondent-level traits as informative, while

interpreting the heterogeneity as suggestive.

6. Underlying Mechanisms

6.1 Possible explanations

Our RCT design seeks to isolate the causal effect of AI bots on managers' willingness to supply information by holding constant the content of investor questions while varying only the managers' perception that the questions are being asked by AI-equipped investors. This allows us to cleanly identify whether managers respond differently when they believe they are addressing an AI-augmented audience. However, while it is evidence that managers reduce disclosure, the mechanism is not uniquely identified because managers can interpret AI both as: (1) enhanced monitoring that raises the expected cost of strategic narrative management (Sims 2003, 2006; Blankespoor et al. 2020; Kim and Kim 2025), and (2) a non-human "alien" presence that induces discomfort or distrust (e.g., Luo et al. 2019; Tong et al. 2021; Yin et al. 2024). Under the first view, which is more about the positive effect of AI, managers may withhold or neutralize disclosures due to the increased risk of being penalized by AI-equipped investors. Under the second and negative view, the psychological uncertainty can lead to distrust, discomfort, or caution, reducing managers' willingness to engage in communication.

In the cross-sectional tests, we show SOEs have an attenuated decline of response likelihood, and greater decline for female and older managers and managers with a computer science background. These results are more consistent with the positive view. We also note that the positive view is more empirically tractable. First, we can empirically examine shifts in the managerial narrative—a key proxy for disclosure management (Huang et al. 2014). Second, we can examine the length and timeliness of managerial responses to AI-styled questions relative to responses to the human-styled counterparts, which allows us to separate the positive view from the negative

one. Third, the real-time format of OECs allows us to investigate the informativeness in managerial responses to AI-styled questions via an intraday event-study design (Baik et al. 2025).

To provide a simple conceptual explanation, we use a stylized framework of strategic disclosure with potential manipulation. This framework is intentionally simplified and serves only as an organizational tool to explain our findings. The goal of this framework is not to fully capture these complexities, but rather to provide a baseline structure that aligns with the logic of our experimental design and explain the full set of findings.[37]

Suppose the manager observes a private value z, and chooses whether to disclose. Disclosure follows a threshold strategy at a disclosure cost c (Verrecchia 1983; Dye 1985; Rogers et al. 2011): disclose happens if and only if $z \geq t$. If the manager discloses, she engages in narrative-based inflation by reporting $t + \delta$, where $\delta \geq 0$. The cost of inflation is quadratic: $\frac{1}{2}\delta^2$.

The manager believes that investors' interpretation of the disclosure depends on whether they use AI. With probability $q \in (0,1)$, investors simply take the price at $t + \delta$. This represents the fact that investors cannot make accurate inferences and are misled by the manager's narrative. With probability $1 - q$, investors see through narrative management and take the price at t. This represents the case in which investors can make accurate inferences. We depart from classical fully rational expectations models of disclosure by assuming that investors process disclosures with bounded inferences. Specifically, investors treat disclosures as either point-valued or transparent with fixed probabilities. These beliefs do not internalize the strategy of the manager and thus prevents unraveling. Managers, anticipating this structured misperception, provide a strategic

[37] We can extend this framework to embrace the richer theoretical foundations of Bertomeu et al. (2025b) and Kim and Kim (2025) and offer a more detailed treatment of how different types of information processing failures, such as bounded rationality and inaccurate inference, can shape both the disclosure threshold and the extent of narrative-based manipulation. As shown by Kim and Kim (2025), this would further increase the disclosure threshold.

narrative and disclosure accordingly. Prices reflect the average of these boundedly rational interpretations. The manager's payoff if disclosing t with inflation δ is

$$U(t,\delta) = q(t+\delta) + (1-q)t - \frac{\delta^2}{2} - c = t + q\delta - c - \frac{\delta^2}{2}. \quad (10)$$

We can optimize the manager's disclosure and narrative manipulation decisions separately, despite being chosen simultaneously, because the narrative choice affects only one part of the objective and does so in a smooth, separable way. Specifically, the manager's total payoff from disclosure is the sum of a baseline value (t), a narrative-dependent benefit (which depends only on q and δ), and a cost of manipulation (also only a function of δ) (Rogers et al. 2011). This structure allows us to solve for the optimal narrative independently by maximizing and then plug the resulting optimized value (the "manipulation envelope") back into the disclosure condition. Because the optimal narrative choice does not depend on the signal or the disclosure threshold, this separation is valid and leads to the same outcome as a sequential game in which the narrative is chosen after the disclosure decision. Therefore, we obtain the following equilibria: the manager optimally implements $\delta^* = q$ and discloses if and only if $U(t,\delta) \geq L(t) = E[\, z \mid z \leq t \,]$. The equilibrium threshold solves: $t^* = L(t^*) + c - \frac{1}{2}q^2$.

When q decreases, the right-hand side of the equation increases. Note that $0 < \frac{\partial L(t)}{\partial t} < 1$ for truncated normal distribution, which also ensures only one equilibrium (Bagnoli and Bergstrom 2005). To restore equilibrium, t^* must rise. In other words, when inaccurate inference is reduced, the threshold for disclosure increases. This yields a first prediction, which is tested in Table 6:

a) *When managers perceive that investors are equipped with AI and have lower processing failure probability q, her disclosure threshold t becomes higher, reducing the propensity of disclosure.*[38]

[38] Yin et al. (2024) highlight a potential stigma effect associated with AI, which may discourage managerial

This model also generates a second testable prediction on the narrative of the disclosure:

b) When managers perceive that investors are equipped with AI and have lower processing failure probability q, she reduces the narrative management δ.[39]

In the following sections, we first directly test the prediction on sentiment adjustments. Then, we study managers' response length and response timeliness to proxy for managers' allocated effort among questions to further tease out the positive view. Finally, we analyze real-time market reactions to AI-styled questions and managerial responses. Relying on these three sets of tests, we seek to demonstrate that the reduced managerial disclosure, in our RCT context, is best explained by AI's disciplinary effect on strategic managerial narrative management.

6.2 Managerial sentiment manipulation

To provide support for the AI disciplinary effect mechanism, we test the second prediction of our model: less manipulation in the narrative when managers respond. Figure 7 provides initial univariate evidence. Panel A shows that managers generally deliver a positive sentiment across four questions, consistent with the institutional setting of firms' discretion in response phrasing. Panel B shows that responses to treated questions contain less positive sentiment than responses to the controls, particularly for the questions on future growth outlook (Q2) and current industry performance (Q3). Panel C shows that neutrality rises under treatment, again most strongly for Q2 and Q3. These patterns are consistent with managers' shifting from more positive, spun responses

communication and help explain the reduced response rate. This negative perception can be interpreted as a psychological or cognitive cost of disclosing to AI—a "mental cost"—which enters the equilibrium condition as: $t^* = L(t^*) + c(AI) - \frac{1}{2}q^2$, where $c(AI)$ represents the perceived cost of communicating with an AI. As shown in the equation, this effect independently raises the disclosure threshold, compounding the discouraging impact on narrative management. Yet this account cannot fully explain the observed tonal shifts or the delayed trading response until managerial disclosures are released, both of which suggest managers remain strategically engaged.

[39] This setting also allows us to examine investor reactions upon the arrival of disclosures. Let $F(t)$ denote the probability of non-disclosure. Prior to the response, the price reflects a weighted average of the expected price under response and non-response: $\big(1 - F(t)\big)(t + q^2) + F(t)L(t)$. The price upon the responses arrive is $t + q^2$. Therefore, when q is manipulated downward, the threshold increases, surprise increases when responses occur (because the expectation is less responses), and the observed jump in price can be dominated by this surprise, leading to greater trading activities.

to more-neutral disclosures when they perceive AI readership.

To formally test this prediction, we run the following OLS regressions:

$$Tone_{q,c} = \alpha_j + \alpha_b + \alpha_p + \alpha_q + \gamma \times Treated_c + \theta \times X_{q,c} + \varepsilon_{q,c}, \quad (11)$$

where $Tone_{q,c}$ represents the two sentiment-related measures. The first is Sentiment, which captures the direction of the response along a positive–negative spectrum. It is measured using FinBERT1 embeddings mapped onto question-specific sentiment axes. Higher values indicate more-positive responses. The second is Neutrality, which captures the degree to which responses avoid positive or negative affect. It is defined as the negative absolute cosine similarity of the response relative to the sentiment axis, so that higher values represent a more-neutral or muted narrative. $X_{q,c}$ comprises sentiment predictors from Huang et al. (2014) in addition to a firm-question-level inverse Mills ratio (IMR) derived from Equation (5) in Section 5.2, as is suggested in Heckman (1976), to account for the truncated sample issue.

Table 8 reports the regression results. The analysis sample in this table is all firm-questions with $Respond_{q,c} = 1$ and non-missing controls. Conditional on responding, treated responses exhibit significantly lower sentiment in column (1): about –0.007 relative to the cosine scale of –1 to 1. Compared to the sample mean sentiment of 0.077 (Table 2, Panel B), this represents a 9.1% reduction in positivity. Neutrality increases by about 0.005 in column (3), which corresponds to a 5.7% rise relative to the sample mean neutrality of –0.088.

An alternative explanation for the less-positive narrative in managerial responses to questions about industry peer performance is that managers adopt a more negative stance toward peers to enhance their own firm's relative standing. However, this interpretation does not align with our empirical evidence. First, we observe a decline in response rates to peer-related questions, suggesting reduced engagement rather than strategic contrast. Second, as in Panel A of Figure 7,

peer-related sentiment remains positive on average, contradicting the notion of intentional negativity. Third, responses to AI-styled questions show increased neutrality, not decreased, indicating a shift toward more-balanced rather than biased language. Thus, these patterns are inconsistent with the view that managers are using peer criticism to improve perceptions.

Figures 8 and 9 show that these disciplinary effects are robust across platforms and questions, but they also reveal interesting nuances. A large portion of platforms exhibit negative effects on sentiment (Figure 8) and positive effects on neutrality (Figure 9) of perceived AI bots. The only exception is Eastmoney, which may be due to small sample bias.

6.3 Do managers take AI-styled questions more seriously?

We next examine whether managers take AI-styled questions more seriously. Conditional on responding, *how* managers respond provides a direct measure of attention and effort. Under the heightened-monitoring view, we expect perceived AI readership to increase managers' incentives to be careful, leading to more-substantive (e.g., longer) responses. By contrast, if the treatment primarily triggers anti-AI sentiment or psychological disengagement, managers should devote less attention to AI-styled questions, responding with shorter, less effortful answers.

We directly test for the impact of perceived AI bots on response length in Table 9. We focus on three length measures: # Characters, # Words, and # Sentences. We take the natural logarithms of these three variables to account for highly right-skewed distributions. Results in Table 9, from column (1) to (3), show a consistent increase in response length when AI bots are present, consistent with managers devoting more effort to AI-styled questions relative to the human-authored counterparts. To mitigate the concern that the observed longer answers to AI-styled questions are driven by longer question prompts (see Table 3), in column (4) of Table 9, we additionally regress Response Time on our treatment indicator, and find slightly increased

timeliness of managerial response to AI questioners. Overall, the longer and quicker responses to AI-styled questions are consistent with managers devoting more effort to potentially intensified monitoring from AI-equipped investors.

6.4 Real-time market responses

A unique advantage of our setting is that on some OEC hosting platforms, both investor questions and managerial responses are posed live and are immediately observable to all market participants (i.e., real investors). This real-time disclosure environment allows us to directly link disclosure to AI bots with contemporaneous market reactions. The prediction is straightforward: if the investors believe that managerial disclosures to AI bots are associated with a higher threshold (i.e., less disclosure) and higher informativeness (i.e., less information manipulation), the incidences of disclosure should elicit stronger trading activity (e.g., k-level thinking).

Our sample is restricted to firm-questions during the two two-hour trading periods (9:30–11:30 and 13:00–14:56) of each trading day.[40] We restrict our sample to OECs only on SSE Roadshow, because this is the only platform on which a posted but unanswered question is immediately and reliably observable to the public before a manager responds, which allows us to precisely capture market responses to both the arrival of the online question arrival and the managerial response. We focus on abnormal trading volume (AbnVolume), a standard variable in event-study and microstructure literature that measures deviations in trading activity relative to a historic average, and is defined as trading volume (thousand shares) over the three- or five-minute window less the preceding five-day average of the mean three- or five-minute trading volume over the same half-day two-hour trading periods (9:30–11:30 or 13:00–4:56). Moreover, as all online

[40] Regulators discourage firms from holding OECs during trading hours. (See, for example, https://www.sse.com.cn/lawandrules/sselawsrules2025/stocks/mainipo/c/c_20250516_10779126.shtml.) This strongly drives down observation counts in our market response tests.

Q&As are observable live, we compute AbnVolume around each real-world investor questions or managerial response and average across all these questions or responses. We include this average as a control for contemporaneous shocks from these co-occurrences.

Table 10 presents the regression results. We regress each market outcome on the treatment indicators in a short-window event study design, controlling for industry, hosting date batch, question fixed effects, and contemporaneous shocks. Standard errors are clustered by firm. Event windows span [–1min, 1min] or [–2min, 2min] around arrival of the experimental question in columns (1) and (2) and around the timestamp of the managerial response to our question in columns (3) and (4). Columns (1) and (2) show that trading volume decreases by around 3,609 to 4,793 shares upon arrival of AI-styled questions, which is around 12% relative to sample standard deviations—a sizable decline given the narrow intraday window.[41] There are several reasons trading activity may decline upon the arrival of AI-styled questions. First, investors may update their beliefs about the sophistication of their counterparties, perceiving the presence of AI-equipped investors, and become more concerned about adverse selection, which can deter trading. Second, if investors expect the forthcoming managerial responses to be more informative, the option value of waiting increases, prompting them to delay trading until the response is revealed (McDonald and Siegel 1986; Acharya et al. 2011). As a result, we observe a modest decline in trading volumes.

However, we do not observe significant differences in abnormal trading volume around managerial responses to AI-styled questions in columns (3) and (4). This muted net effect is consistent with two offsetting forces. On the one hand, AI questioners discipline managers and limit manipulation, so the ensuing answer resembles a more-precise public signal that coordinates

[41] The sample standard deviations (untabulated) are 28,000 and 40,847 shares in [–1min, 1min] and [–2min, 2min], respectively.

investors' beliefs and reduces disagreement, which should dampen volume (Morris and Shin 2002; Chang et al. 2022). On the other hand, because AI scrutiny raises the disclosure threshold, managers may answer in detail only when their private information is sufficiently favorable, so conditional on disclosure, the answer conveys a stronger positive or more credible signal, which tends to stimulate trading. The near-zero coefficients in columns (3) and (4) suggest that these disagreement-reducing and threshold-increasing effects cancel out in the narrow response window. We acknowledge that greater informativeness need not always reduce trading: when investors hold heterogeneous prior beliefs, more-informative disclosures can also raise volume through differential belief updating. This is precisely why the net effect on volume around responses is theoretically ambiguous, and our near-zero estimate is consistent with the disagreement-reducing and threshold-increasing forces offsetting each other.

7. Conclusion

This paper provides causal evidence that the perceived presence of AI-equipped investors reshapes corporate disclosure behavior and capital market reactions. Using a large-scale RCT of listed firms in China, we manipulate only the perceived identity of investor questioners, framing them as AI or human, while holding information content constant. This design isolates the strategic effects of AI on managerial behavior, overcoming the dual unobservability of investor AI adoption and managerial belief formation.

We find that managers are significantly less likely to respond when they perceive the presence of AI bots. Linking responses to firm and manager characteristics, we find a smaller reduction in response likelihood among SOEs, and a greater reduction among female managers, older managers and managers with a computer science background.

The observed decline in information production by managers may reflect both the *positive*

role of AI in disciplining strategic disclosure (Kim and Kim 2025) and the *negative role* of AI in imposing a psychological or cognitive cost (Yin et al. 2024). Our findings are more in line with the positive channel. Specifically, narrative analysis reveals that managers adopt a more-neutral narrative when they believe they are talking to AI, suggesting less opportunistic narrative management. Managers also provide longer and timelier responses to AI-styled questions, implying that they treat these questions more carefully. Simultaneously, equity markets tend to be less reactive to AI-styled questions, consistent with investors waiting for the more informative response by managers that they anticipate.

Finally, firms themselves may increasingly rely on AI tools to help prepare their responses (Blankespoor et al. 2026). We cannot rule out such AI-assisted responding, and the interaction between AI-generated questions and AI-assisted answers is an interesting question that we leave to future research.

References

Abramova, I., Core, J., and Sutherland, A. 2020. Institutional investor attention and firm disclosure. *Accounting Review*, 95(6): 1-21.

Acharya, V. V., P. Demarzo, and I. Kremer. 2011. Endogenous information flows and the clustering of announcements. *American Economic Review*, 101(7): 2955-2979.

Alesina, A., S. Stantcheva, and E. Teso. 2018. Intergenerational mobility and preferences for redistribution. *American Economic Review,* 108(2): 521-554**.**

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

Bagnoli, M., and T. Bergstrom. 2005. Log-concave probability and its application. *Economic Theory*, 26: 445-469.

Baik, B., A. G. Kim, D. S. Kim, and S. Yoon. 2025. Vocal delivery quality in earnings conference calls. *Journal of Accounting and Economics*, forthcoming.

Benjamini, Y, and D. Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29: 1165-1188.

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

Bertomeu, J., Y. Lin, Y. Liu, and Z. Ni. 2025b. Voluntary disclosure, misinformation, and AI information processing: Theory and evidence. *SSRN Working Paper*.

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther. 2010. The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, 50(2-3): 296-343.

Bird, A., and S. A. Karolyi. 2016. Do institutional investors demand public disclosures? *Review of Financial Studies*, 29(12): 3245-3277.

Blankespoor, E. 2022. Understanding investor interaction with firm information: A discussion of Lee and Zhong (2022). *Journal of Accounting and Economics*, 74: 101523.

Blankespoor, E., E. DeHaan, and I. Marinovic. 2020. Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics*, 70(2-3).

Blankespoor, E., E. DeHaan, and Q. Li. 2026. "Generative AI in Financial Reporting." *Journal of Accounting Research* (March).

Boone, A. L., and J. T. White. 2015. The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics*, 117(3): 508-533.

Bradshaw, M. T., C. Ma, B. P. Yost, and Y. Zou. 2025. Generative AI use by market information intermediaries: Evidence from Seeking Alpha. *Working Paper*.

Brislin, R. W. 1970. Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3): 185-216.

Camerer, C. F., T.-H. Ho, and J.-K Chong. 2004. A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3): 861-898.

Cao, S., W. Jiang, B. Yang, and A. L. Zhang. 2023. How to talk when a machine is listening: Corporate disclosure in the age of AI. *Review of Financial Studies*, 36(9): 3603-3642.

Cao, S., W. Jiang, J. Wang, and B. Yang. 2024. From man vs. machine to man + machine: The art and AI of stock analyses. *Journal of Financial Economics*, 160: 103910.

Chan, T. Y., L. Liao, X. Martin, and Z. Wang. 2024. Avoiding peer information and its effects on charity crowdfunding: A field experiment. *Management Science*, 70(4): 2272-2293.

Chang, A., X. Dong, X. Martin, and C. Zhou. 2025. AI democratization and trading inequality. *SSRN Working Paper*.

Chang, Y., P. Hsiao, A. Ljungqvist, and K. Tseng. 2022. Testing disagreement models. *Journal of Finance*, 67(4): 2239-85.

Chetty, R. 2015. Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, 105(5): 1-33.

Cheng, Q., P. Lin, and Y. Zhao. 2025. Does generative AI facilitate investor trading? Early evidence from ChatGPT outages. *Journal of Accounting and Economics*, forthcoming.

Costello, Anna M., Andrea K. Down, and Mihir N. Mehta. 2020. "Machine + Man: A Field Experiment on the Role of Discretion in Augmenting AI-Based Lending Models." *Journal of Accounting and Economics* 70(2–3):101360.

Croom, Joe. 2025. Interactivity and illusions of ability: How using generative AI affects investor judgments. *Journal of Accounting Research*, forthcoming.

Dong, Y., O. Z. Li, Y. Lin, and C. Ni. 2016. Does information-processing cost affect firm-specific information acquisition? Evidence from XBRL adoption. *Journal of Financial and Quantitative Analysis*, 51(2): 435-462.

Dye, R. A. 1985. Disclosure of nonproprietary information. *Journal of Accounting Research*, 23: 123-145.

DellaVigna, S., and E. Linos. 2022. RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica*, 90(1): 81-116.

Duflo, E., R. Glennerster, and M. Kremer. 2007. Chapter 61 Using randomization in development economics research: A toolkit. *Handbook of Development Economics*, 4(7): 3895-3962.

eToro.com. 2025. Retail investors flock to AI tools, with usage up 46% in one year. https://www.etoro.com/news-and-analysis/etoro-updates/retail-investors-flock-to-ai-tools-with-usage-up-46-in-one-year/.

Gao, P. 2008. Keynesian beauty contest, accounting disclosure, and market efficiency. *Journal of Accounting Research*, 46(4): 785-807.

Gao, P., and P. J. Liang. 2013. Informational feedback, adverse selection, and optimal disclosure policy. *Journal of Accounting Research*, 51(5), 1133-1158.

Grossman, S. J. 1981. An introduction to the theory of rational expectations under asymmetric information. *Review of Economic Studies*, 48(4): 541-559.

Hambrick, D. C., and P. A. Mason. 1984. Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review*, 9(2): 193-206.

He, G., and Q. Xiahou. 2025. The demand, supply, and market responses of corporate ESG actions: Evidence from a nationwide field experiment in China. *Working Paper*.

Healy, M. P., and K. G. Palepu. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, 31(1-3): 405-440.

Heckman, J. J. 1976. The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5: 475-492.

Huang, A. H., H. Wang, and Y. Yang. 2023. FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40(2): 806-841.

Huang, J., and D. J. Kisgen. 2013. Gender and corporate finance: Are male executives overconfident relative to female executives? *Journal of Financial Economics*, 108(3): 822-839.

Huang, X., S. H. Teoh, and Y. Zhang. 2014. Tone management. *Accounting Review*, 89(3): 1083-1113.
Jha, M., H. Liu, and A. Manela. 2025. Does finance benefit society? A language embedding approach. *Review of Financial Studies*, 0(0): 1-40.
Jia, N., N. Li, G. Ma, and D. Xu. 2025. Corporate responses to generative AI: Early evidence from conference calls. *Review of Accounting Studies*.
Jones, C. M., D. Shi, X. Zhang, and X. Zhang. 2020. Heterogeneity in retail investors: Evidence from comprehensive account-level trading and holding data. *Working Paper*.
Jung, W.-O., and Y. K. Kwon. Disclosure when the market is unsure of information endowment of managers. *Journal of Accounting Research*, 26(1): 146-153.
Kim, J. M., and C. Kim. 2025. Different types of information processing costs and voluntary disclosure. *Working Paper*.
Klotz, A. C., B. W. Swider, S. H. Kwon. 2023. Back-translation practices in organizational research: Avoiding loss in translations. *Journal of Applied Psychology*, 2023, 108(5): 699-727.
Lee, C. M.C., and Q. Zhong. 2022. Shall we talk? The role of interactive investor platforms in corporate communication. *Journal of Accounting and Economics*, 74: 101524.
Liang, P. J., and G. Zhang. 2019. On the social value of accounting objectivity in financial stability. *Accounting Review*, 94(1): 229-248.
Liao, L., Martin, X., Wang, N., Wang, Z., and Yang, J. 2023. What if borrowers were informed about credit reporting? Two natural field experiments. *Accounting Review*, 98(3): 397-425.
List, J. A. 2011. Why economists should conduct field experiments and 14 tips for pulling one off. *Journal of Economic Perspectives*, 25(3): 3-16.
Lu, Y., H. Shi, and X. Zhou. 2025. Impact of digital technology risk exposure on corporate value in Chinese firms: Evidence from text analysis using large language models. *Economic Research Journal*, (2): 73-89.
Luo, X., S. Tong, Z. Fang, and Z. Qu. 2019. Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, 38(6): 937-947.
Mayew, W. J., M. Sethuraman, and M. Venkatachalam. 2020. Individual analysts' stock recommendations, earnings forecasts, and the informativeness of conference call question and answer sessions. *Accounting Review*, 95(6): 311-337.
McCroskey, J. C., and J. J. Teven. 1999. Goodwill: A reexamination of the construct and its measurement. *Communications Monographs*, 66(1): 90-103.
McDonald, R., and Siegel, D. 1986. The value of waiting to invest. *Quarterly Journal of Economics*, 101(4), 707-727.
Mercer. 2025. AI integration in investment management. https://www.mercer.com/assets/global/en/shared-assets/global/attachments/pdf-2024-Mercer-AI-integration-in-investment-management-2024-global-manager-survey-report-03212024.pdf.
Morris, S., and H. S. Shin. 2002. Social value of public information. *American Economic Review*, 92(5): 1521-34.
Nagel, R. 1995. Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5), 1313-1326.
Rajan, R., P. Ramella, and L. Zingales. 2023. What purposes do corporations support? Evidence from letters to shareholders. *NBER Working Paper*: 31054.
Rennekamp, K. M., M. Sethuraman, and B. A. Steenhoven. 2022. Engagement in earnings

conference calls. *Journal of Accounting and Economics*, 74(1), 101498.
Rogers, J. L., A. V. Buskirk, and S. L. C. Zechman. 2011. Disclosure tone and shareholder litigation. *Accounting Review*, 86(6): 2155-2183.
Russell, S. J., and P. Norvig. 1995. Artificial intelligence: A modern approach. *Englewood Cliffs*, NJ: Prentice Hall.
Serfling, M. A. 2014. CEO age and the riskiness of corporate policies. *Journal of Corporate Finance*, 25: 251-273.
Sheng, J., Z. Sun, B. Yang, and A. L. Zhang. 2025. Generative AI and asset management. *SSRN Working Paper*.
Shleifer, A., and R. W. Vishny. 1994. Politicians and firms. *Quarterly Journal of Economics* 109(4): 995-1025.
Sims, C. A. 2003. Implications of rational inattention. *Journal of Monetary Economics*, 50(3): 665-690.
Sims, C. A. 2006. Rational inattention: Beyond the linear-quadratic case. *American Economic Review*, *96*(2): 158-163.
Tong, S., N. Jia, X. Luo, and Z. Fang. 2021. The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, 42(9): 1600-1631.
Tsinghua PBCSF, and Ant Group Research. Research on current situation and future trend of AI adoption in wealth management services. https://www.pbcsf.tsinghua.edu.cn/AIcaifuguanlifuwuxianzhuangyuqushi.pdf.
Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science*, 211(4481), 453-458.
Verrecchia, R. E. 1983. Discretionary disclosure. *Journal of Accounting and Economics*, 5: 179-194.
Wong, T.J., Y. Yi, G. Yu, S. Zhang, and T. Zhang. 2025. Enhancing investor engagement with AI-summarized disclosures. *Working Paper*.
Wong, T.J., G. Yu, S. Zhang, and T. Zhang. 2024. Calling for transparency: Evidence from a field experiment. *Journal of Accounting and Economics*, 77(1), 101604.
Yin, Y., N. Jia, and C. J. Wakslak. 2024. AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences of the United States of America*, 121(14), art. no. e2319112121.
Zhang, T., K. Tang, and H. Xie. 2025. Construction and macroeconomic forecasting of the chief economist confidence indexes based on text analysis. *Journal of Management Sciences in China*, 28(3): 1-18.

Variable Definitions

Panel A. Firm (conference) level	
Variable name	Definition
Treated	1 for treatment group OECs, and 0 for control group OECs.
Text	Simply equivalence of Treated, as each treatment group OEC receives the identical set of AI-styled questions.
Image	1 for OECs with Treated=1 in which we raise questions with an explicit AI label (i.e., a robotic profile photo), and 0 otherwise.
Age	Decimal number of years between April 14, 2025, and establishment date. The establishment date data is from the Chinese Stock Market and Accounting Research (CSMAR) database.
Size	Natural logarithm of market capitalization snapshot on December 31, 2024. The stock trading data are from CSMAR.
Book to Market	Ratio of book equity to market value snapshot on December 31, 2024. The financial and stock trading data are from CSMAR.
Leverage	Debt to asset ratio at 2024 fiscal end. The financial data are from CSMAR.
Annual Return	Annual return over 2024. Directly derived from CSMAR.
Return Volatility	The standard deviation of monthly returns over 2024. The stock trading data are from CSMAR.
ROA	Return on assets over 2024 fiscal year. The financial data are from CSMAR.
Performance Volatility	The standard deviation of annual returns on assets over the latest five years, with at least three years of data available. The financial data are from CSMAR.
Loss	1 if ROA is less than 0, and 0 otherwise.
ΔROA	Return on assets over 2024 fiscal year minus return on assets over 2023 fiscal year. The financial data are from CSMAR.
EPS Rise	1 if earnings per share over 2024 fiscal year is greater than earnings per share over 2023 fiscal year. The financial data are from CSMAR.
Abs. ΔEPS	Absolute value of earnings per share over 2024 minus earnings per share over 2023, scaled by close price on December 31, 2024. The financial and stock trading data are from CSMAR.
Earnings Surprise	Earnings per share over 2024 fiscal year minus Analyst Forecast, scaled by close price on December 31, 2024. The financial and stock trading data are from CSMAR.
Analyst Forecast	The median of analyst forecasts of earnings per share issued during 2024 for the 2024 fiscal year. For firms with no analysts following during 2024, we alternatively use realized earnings per share over 2023, assuming that analysts follow a naïve forecast scheme. The analyst forecast data and financial data are from CSMAR.
Analysts Following	Natural logarithm of 1 plus total number of unique analysts following during 2024. The analyst forecast data are from CSMAR.
Inst. Ownership	Proportion of shares held by institutional investors in all outstanding shares. The ownership structure data is from CSMAR.
Board Independence	Proportion of independent directors in all board members. The board composition data are from CSMAR.
R&D Intensity	Ratio of R&D expenditure to revenues. The financial data are from CSMAR.
SOE	1 for state-owned enterprises, and 0 otherwise. The ownership structure data are from CSMAR.
Customer Concentration	Proportion of sales to firm's top 5 customers. The data are from CSMAR.
Supplier Concentration	Proportion of purchases from firm's top 5 suppliers. The data are from CSMAR.
Elderly	1 if the responding manager's age is among the top 25% of sample, and 0 otherwise. The data are from CSMAR.
Female	1 if the responding manager is female, and 0 otherwise. The data are from CSMAR.
Computer Science	1 if the responding manager holds a computer science degree, and 0 otherwise.

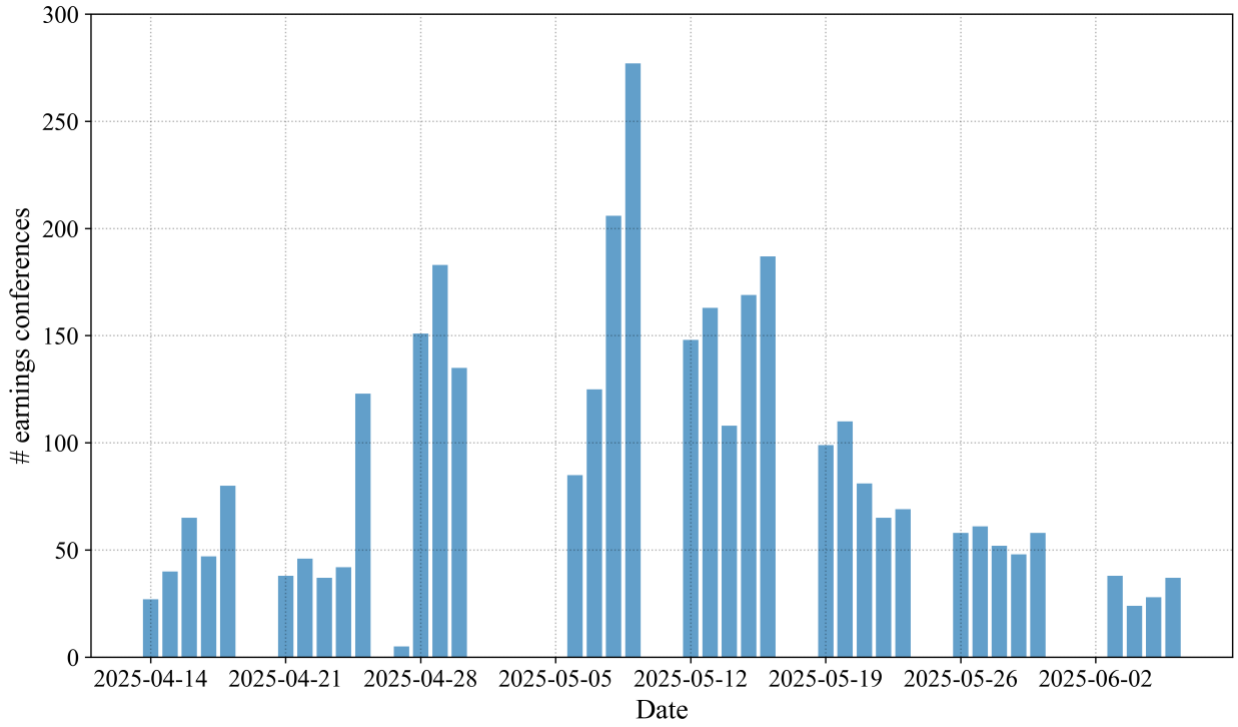
	The data are from CSMAR.
# Q&As	The total count of textual Q&As during OEC, excluding Q&As initiated by our experimenters.
Duration	Duration of an OEC, in minutes.
# Responses	Number of experimental questions responded to.
Non-Response	1 if the firm responds to none of our questions, and 0 otherwise.

Panel B. Firm-question level

Variable name	Definition
Response	1 if the experimental question gets answered, and 0 otherwise.
Sentiment	For question q, we first define a sentiment axis as follow, $$Axis_q = \text{avg}(\{\text{emb}(s): s \in \mathcal{P}_q^+\}) - \text{avg}(\{\text{emb}(s): s \in \mathcal{P}_q^-\}),$$ where $\text{emb}(\cdot)$ is a FinBERT1 transformation that embeds sentences to a 768-dimension vector, $\mathcal{P}_q^+$ and $\mathcal{P}_q^-$ are the predefined sets of positive and negative benchmark responses to question q, respectively, and $\text{avg}(\cdot)$ is a function to compute an equal-weighted average vector of a vector set. Appendix C provides a full list of predefined benchmark responses to each of our four experimental questions. Then, the sentiment score for a response to question q on conference c is computed as $$Sentiment_{q,c} = \frac{1}{N_{q,c}} \sum_s \cos(\text{emb}(s), Axis_q) = \frac{1}{N_{q,c}} \sum_s \frac{\text{emb}(s) \times Axis_q}{|\text{emb}(s)| \times |Axis_q|},$$ where each sentence s is parsed from this response based on the occurrences of periods, exclamation points, and question marks, and $N_{q,c}$ is the count of parsed sentences from this response.
Neutrality	For a management response to question q on conference c, Neutrality is computed as $$Neutrality_{q,c} = -\frac{1}{N_{q,c}} \sum_s |\cos(\text{emb}(s), Axis_q)|,$$ where each sentence s is parsed from this response based on the occurrences of periods, exclamation points, and question marks, and $N_{q,c}$ is the count of parsed sentences from this response.
AbnVolume	Abnormal trading volume (in thousands) over a specific 3- or 5-minute intraday window around either the timestamp when we post our question or the timestamp when a manager responds to our question. The benchmark volume is the preceding 5-day average of mean 3- or 5-minute trading volumes over the same half-day 2-hour trading periods as the OEC.
Response Time	The total minutes management takes to respond to our experimental question.
# Characters	Count of Chinese characters of managerial response to our experimental question, including punctuation.
# Words	Count of words of managerial response to our experimental question, excluding punctuation and stop-words. Words are identified via the *jiaba* algorithm, a pretrained Chinese text parser.
# Sentences	Count of sentences of managerial response to our experimental question. A response is cut into sentences based on the occurrences of periods, exclamation points, question marks, and colons.

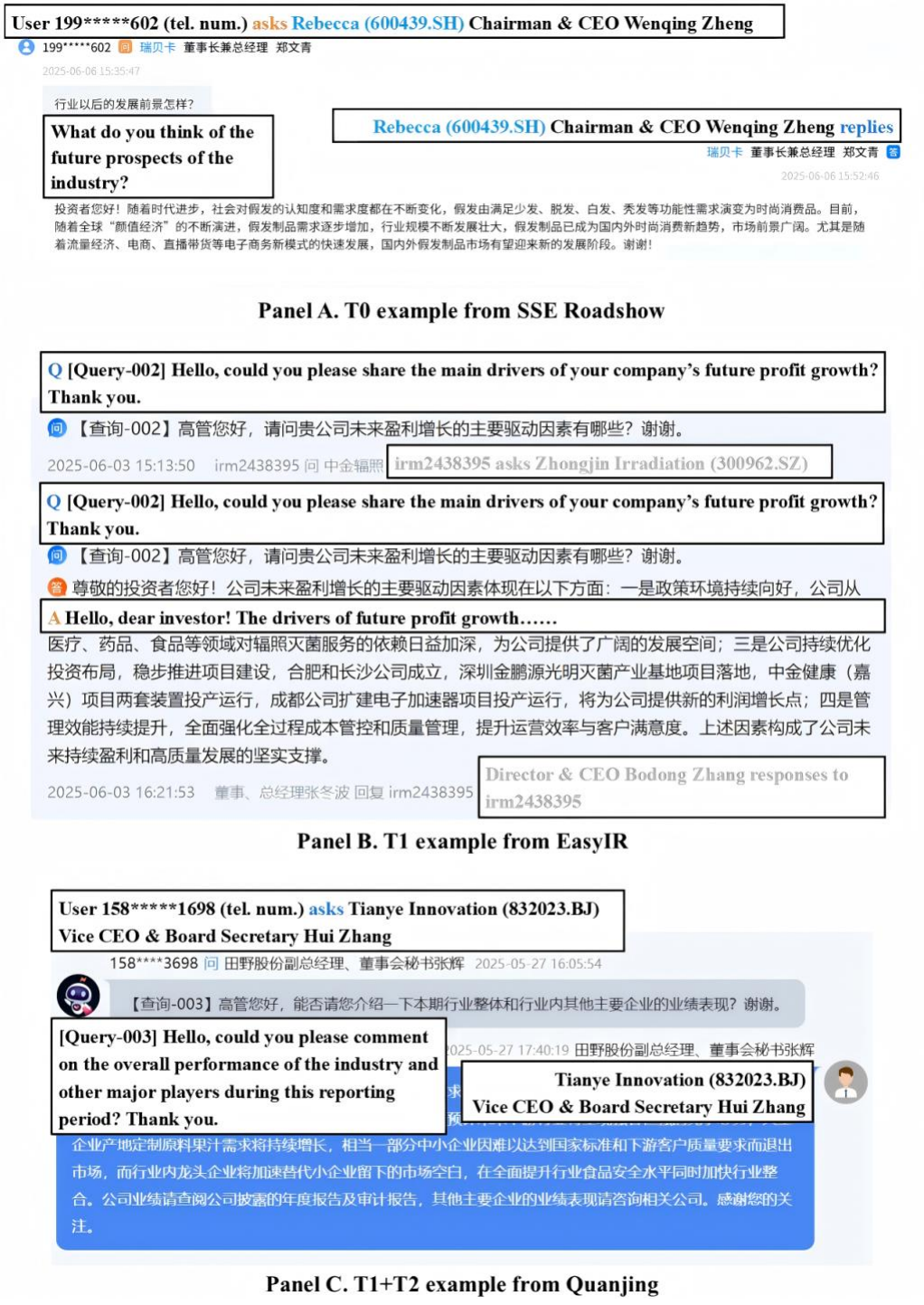
Figures and Tables

Figure 1. Hosting Date Distribution of Announced Earnings Conferences



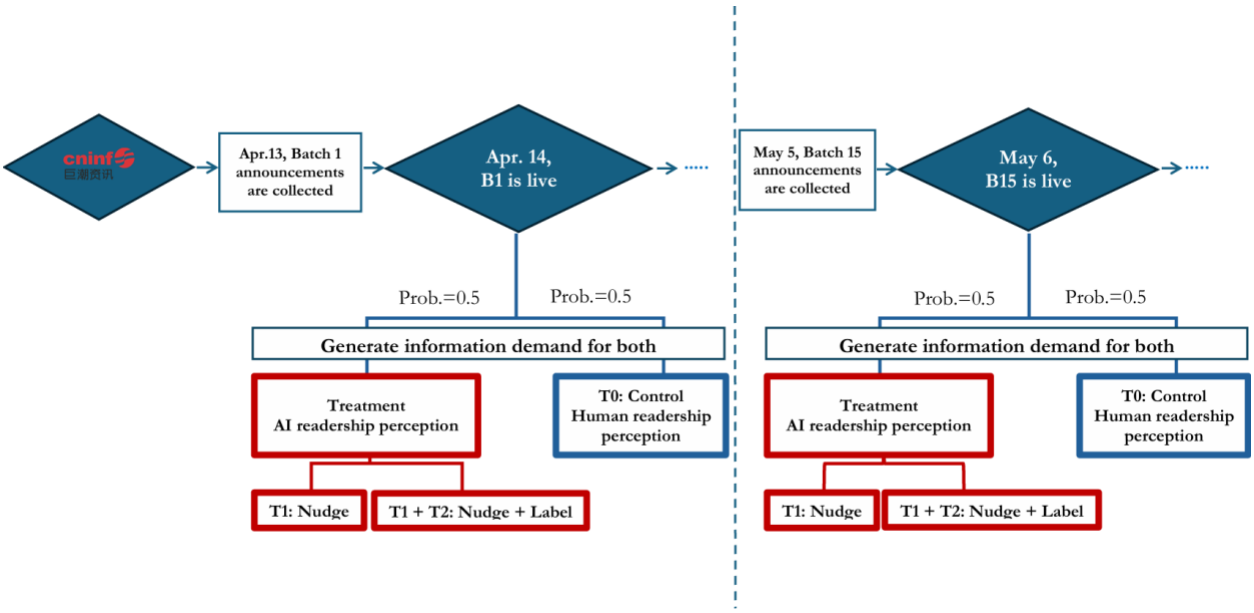
This figure presents the date distribution of earnings conferences within our sample period, April 14 to June 6, 2025. Blue bars indicate daily counts of hosted conferences.

Figure 2. Examples of Experimental Questions and Responses



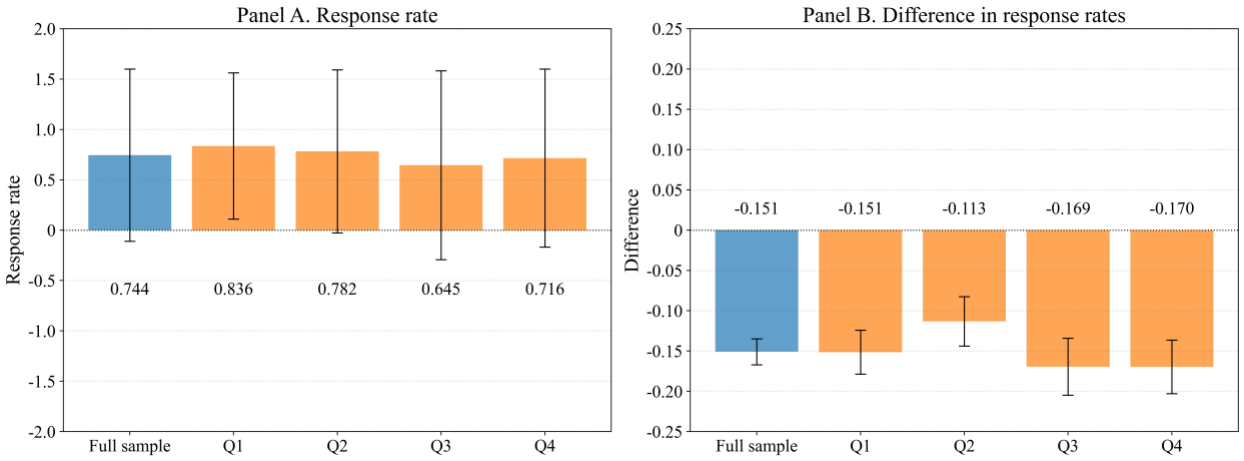
This figure provides examples of treatment arms and their real-world management responses. Panel A is an example of control group question (T0) from SSE Roadshow; Panel B is an example of AI-styled question (T1) from EasyIR; Panel C is an example of AI-styled question with robotic profile visual (T1+T2) from Quanjing. In all panels, English translation is provided for the ease of understanding.

Figure 3. Experiment Design



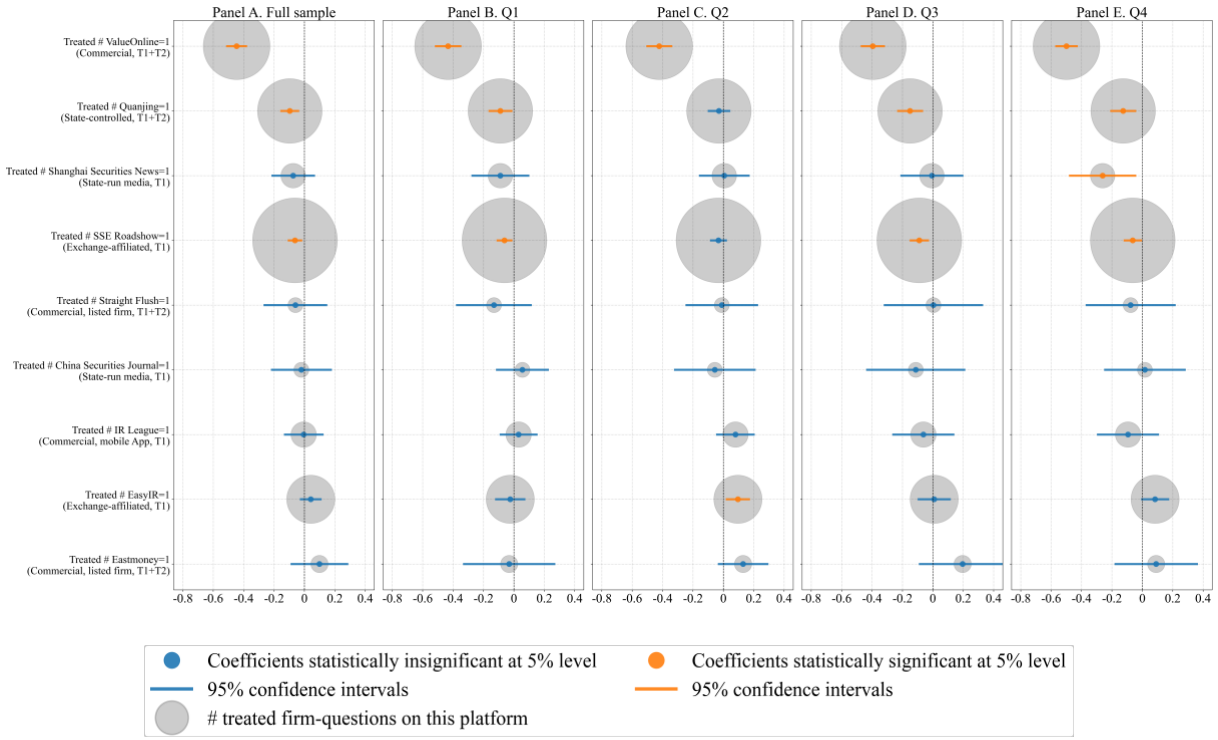
This figure outlines our experiment protocol. We script all earnings conference announcements from the cninf website, and keep the hosting date of each announced earnings conference. Each OEC after sample screening is naturally grouped into hosting date batches and receives each of the two treatments with equal probability on its presentation date.

Figure 4. Managers' Likelihood of Responding



This figure summarizes response rate of our experimental questions both in the full sample (blue bar) and by question (orange bars); Q1, Q2, Q3, and Q4 refer to experimental question 1, 2, 3 and 4, respectively. In Panel A, bars indicate response rates, and black error bars indicate 95% confidence intervals of the response rate. In Panel B, we conduct univariate comparisons of response rates between the treatment and control groups; bars indicate differences in means, and black error bars indicate 95% confidence intervals of the differences derived based on Welch's adjusted standard errors.

Figure 5. Likelihood of Responding by Platform and Question

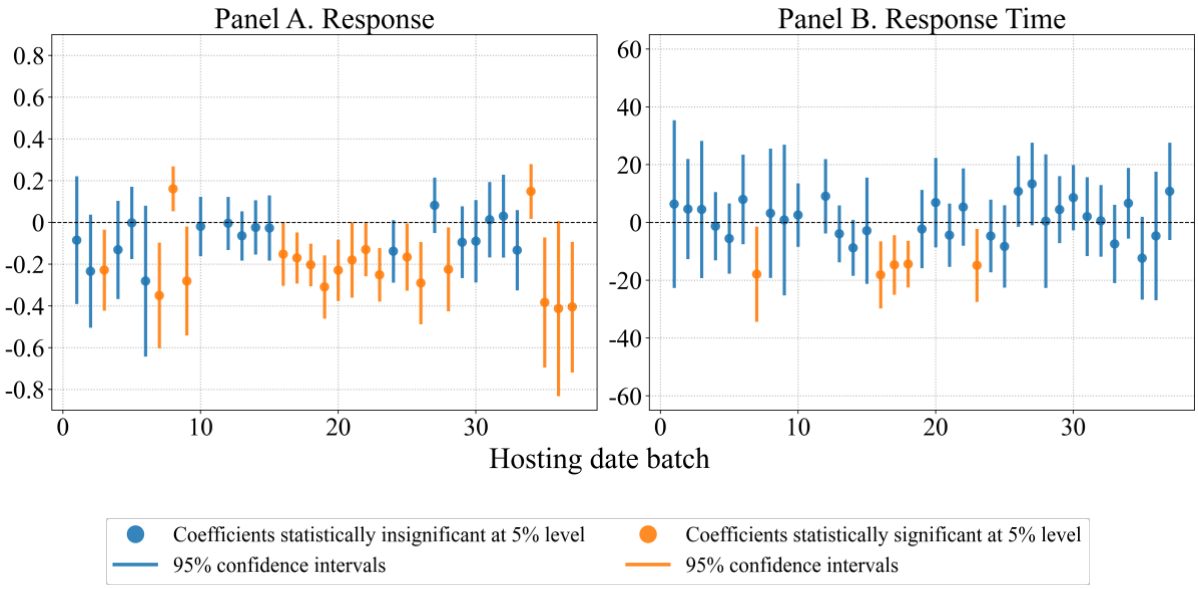


In this figure, we run the following Probit estimations,

$$Response_{q,c} = \alpha_j + \alpha_b + \alpha_p + \alpha_q + \beta_p \times \sum_p Treated_c \times 1(Platform_c = p) + \theta \times X_c + \varepsilon_{q,c}, \quad (7)$$

where q denotes question, c denotes conference (firm), j denotes industry, b denotes conference date batch, and p denotes platform; α_j, α_b, α_p, and α_q are industry, batch, platform, and question fixed effects, respectively; $Respond_{q,c}$ equals 1 if this firm-question receives a reply, and 0 otherwise; $Treated_c$ equals 1 for treated firm-questions, and 0 otherwise; $1(\cdot)$ is an indicator function, and $Platform_c$ is the hosting platform of conference c; X_c is firm-level control variables including Size, Book to Market, Leverage, Annual Return, Return Volatility, Loss, EPS Rise, Abs. ΔEPS, Analysts Following, Inst. Ownership, Board Independence, R&D Intensity, SOE, Customer Concentration, and Supplier Concentration. Table Variable Definitions provides detailed variable definitions. In Panel A, we pool all firm-question observations and cluster standard errors by firm, and through Panels B to E, we estimate respectively for each of our four questions using robust standard errors. This figure plots average marginal effects (blue and orange dots) and their 95% confidence intervals (error bars). Additionally, we outline, both in the full sample and within each question, the number of firm-questions from the treatment group on each platform via the sizes of grey solid circles.

Figure 6. Cross-Firm Contamination



In Panel A of this figure, we estimate the following Probit model,

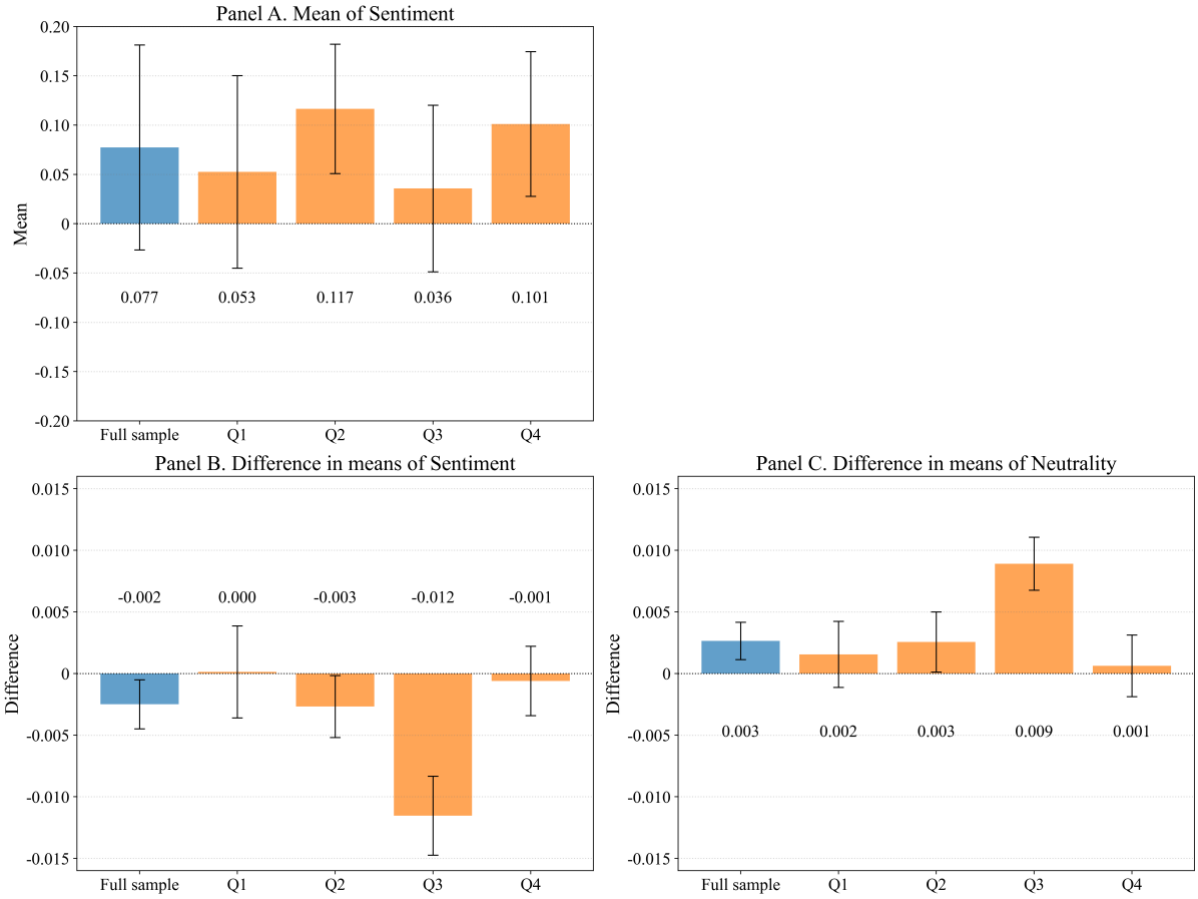
$$Response_{q,c} = \alpha_j + \alpha_b + \alpha_p + \alpha_q + \sum_{b=1}^{37} \beta_b \times Treated_c \times 1(Batch_c = b) + \theta X_c + \varepsilon_{q,c}. \quad (8)$$

In Panel B, we run the following Heckman's (1976) second-stage OLS regression,

$$Response\ Time_{q,c} = \alpha_j + \alpha_b + \alpha_p + \alpha_q + \sum_{b=1}^{37} \beta_b \times Treated_c \times 1(Batch_c = b) + IMR_{q,c} + \varepsilon_{q,c}. \quad (9)$$

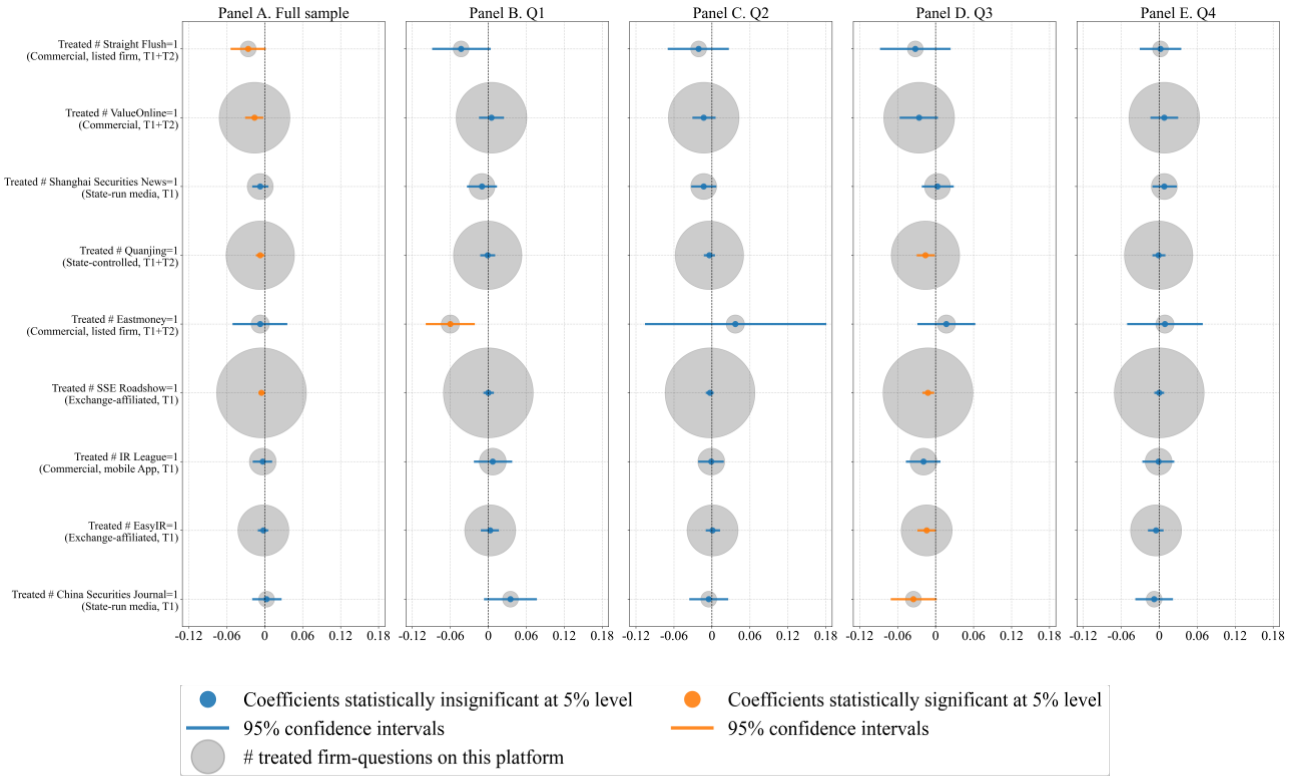
In both panels, q denotes question, c denotes conference (firm), and b denotes conference date batch. α_j, α_b, α_p, and α_q are industry, batch, platform, and question fixed effects, respectively. $Treated_c$ equals 1 for treatment group OECs, and 0 otherwise; $1(\cdot)$ is indicator function, and $Batch_c$ is the conference date batch of conference c; X_c is firm-level control variables including Size, Book to Market, Leverage, Annual Return, Return Volatility, Loss, EPS Rise, Abs. ΔEPS, Analysts Following, Inst. Ownership, Board Independence, R&D Intensity, SOE, Customer Concentration, and Supplier Concentration; $IMR_{q,c}$ in Equation (9) is the inverse Mills ratio derived from Equation (8). See Table Variable Definitions for detailed variable definitions. Our experiment encompasses a total of 37 distinct and potentially non-continuous calendar days, resulting in 37 hosting date batches. We have few observations from batch 11 (i.e., April 27, 2025), therefore, we omit estimation for β_{11} in both panels. Panel A plots the average marginal effects and their 95% intervals; Panel B shows the regression estimates of β_b and their 95% confidence intervals, over different values of b from 1 to 37.

Figure 7. Managerial Sentiment Manipulation



This figure provides a univariate summary of our narrative measures both in the full sample (blue bar) and by question (orange bars); Q1, Q2, Q3, and Q4 refer to experimental question 1, 2, 3, and 4, respectively. In Panel A, bars indicate means of Sentiment, and black error bars indicate 95% confidence intervals of means. In Panels B and C, we conduct univariate comparisons of means of Sentiment and Neutrality, respectively, between the treatment and control groups. Bars indicate differences in means, and black error bars indicate 95% confidence intervals of differences derived based on Welch's adjusted standard errors. Table Variable Definitions provides detailed variable definitions.

Figure 8. Sentiment by Platform and Question

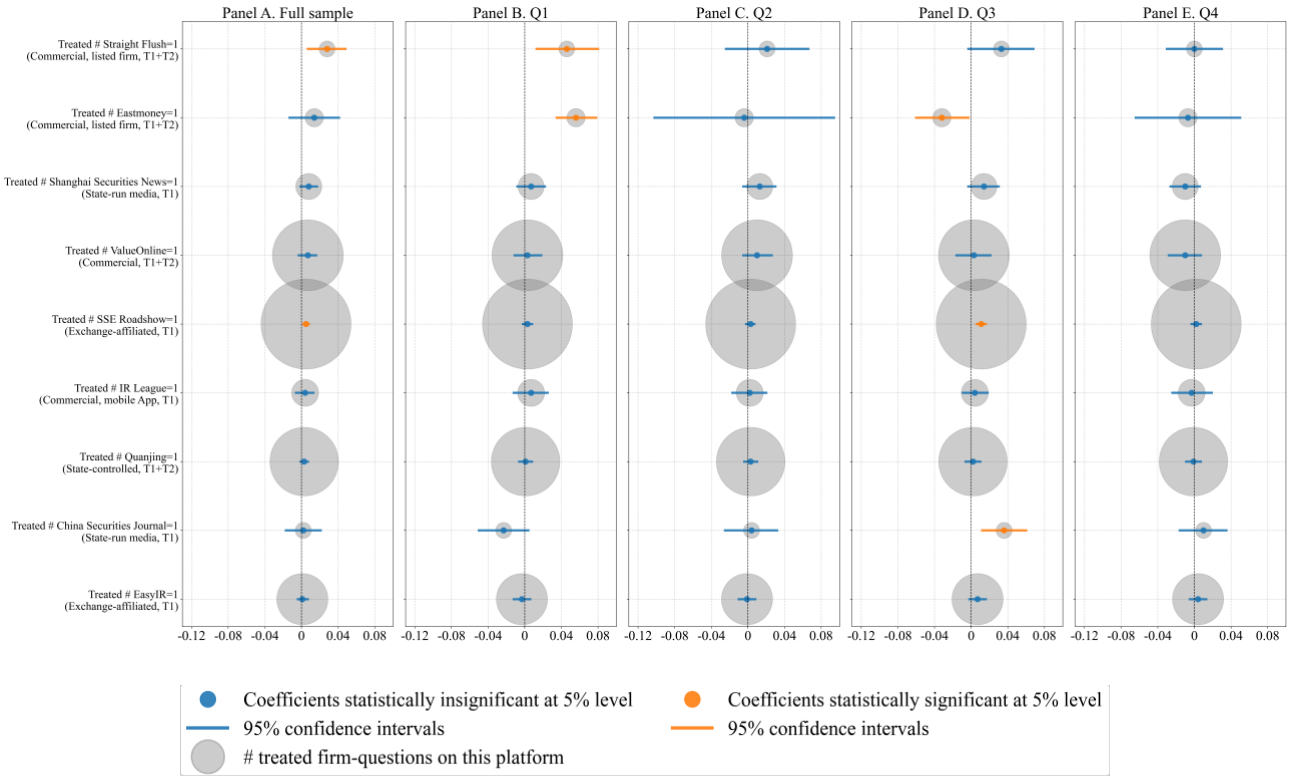


This figure reports the results of the following Heckman's (1976) second-stage OLS regressions,

$$Sentiment_{q,c} = \alpha_j + \alpha_b + \alpha_p + \alpha_q + \beta_p \times \sum_p Treated_c \times 1(Platform_c = p) + \theta \times X_{q,c} + \varepsilon_{q,c}, \quad (12)$$

where $Sentiment_{q,c}$ is the sentiment measure of management response q in OEC c, and $X_{q,c}$ is firm-question level IMR derived from Equation (7) in addition to all firm-level sentiment predictors used in Huang et al. (2014): Age, Size, Book to Market, Annual Return, Return Volatility, ROA, Performance Volatility, Loss, ΔROA Earnings Surprise, and Analyst Forecast. All other specifications remain as in Equation (7). Table Variable Definitions provides detailed variable definitions. In Panel A, we pool all firm-question observations and cluster standard errors by firm; through Panels B to Panel E, we conduct estimation respectively for each of our four questions using robust standard errors. This figure plots estimates of β_p (blue and orange dots) and their 95% confidence intervals (error bars). Additionally, we outline, both in the full sample and within each question, the number of firm-questions from the treatment group on each platform via sizes of the grey solid circles.

Figure 9. Neutrality by Platform and Question



This figure reports the results of the following Heckman (1976) second-stage OLS regressions,

$$Neutrality_{q,c} = \alpha_j + \alpha_b + \alpha_p + \alpha_q + \beta_p \times \sum_p Treated_c \times 1(Platform_c = p) + \theta \times X_{q,c} + \varepsilon_{q,c}, \quad (13)$$

where $Neutrality_{q,c}$ is the sentiment neutrality measure of management response q in OEC c. All other specifications remain as in Equation (12). Table Variable Definitions provides detailed variable definitions. In Panel A, we pool all firm-question observations and cluster standard errors by firm; through Panels B to E, we conduct estimation respectively for each of our four questions using robust standard errors. This figure plots estimates of β_p (blue and orange dots) and their 95% confidence intervals (error bars). Additionally, we outline, both in the full sample and within each question, the number of firm-questions from the treatment group on each platform via the sizes of grey solid circles.

Table 1. Sample Screening

Starting point	3481
1. Less non-OECs	(33)
	3448
2. Less duplicate OECs	(52)
	3396
3. Less JECs	(58)
	3338
4. Less OECs held on platforms that screen investors	(67)
	3271
5. Less pseudo OECs	(5)
	3266
6. Less quarterly report OECs	(108)
	3158
7. Less uncontrollable external factors	(436)
Final sample	2722

This table outlines our sample screening procedure. We start with a total of 3,481 earnings conferences announced to be hosted between April 14 and June 6, 2025, which is our sample period. First, we exclude non-OECs. Second, we exclude duplicate OECs of firms who hosted multiple OECs within our sample period. Third, we exclude joint earnings conferences (JECs) hosted by multiple firms. Fourth, we exclude OECs on platforms that screen participants. Fifth, we exclude OECs with an offline venue in which firms actually do not respond to online queries. Sixth, we exclude OECs hosted separately for quarterly financial reports. Finally, we exclude OECs for which we were unable to conduct the experiments due to external factors.

Table 2. Platform Outline

Platform	Treatment level	Notable features	# (%)
SSE Roadshow	**T0/T1:** Sign up with phone number, and phone number serves as username. No profile customization allowed.	Affiliated with the Shanghai Stock Exchange. A specialized channel for both real time text-based communication (including OEC) and live broadcast of on-site corporate event. Provided freely to listed firms for OEC.	1051 (38.6%)
ValueOnline	**T0/T1+T2:** Allows for both user-uploaded profile image and customized username.	Commercial service. Provide solutions for a variety of corporate communications including OEC.	575 (21.1%)
Quanjing	**T0/T1+T2:** Allows for both user-uploaded profile image and customized username.	State-controlled commercial service. Controlled by Securities Times and ultimately controlled by People's Daily.	561 (20.6%)
EasyIR's Cloud Interview	**T0/T1:** System-assigned username with uniform format (6 digits following "irm," such as "irm2396418"). No profile customization allowed.	EasyIR is an interactive investor platform established by the Shenzhen Stock Exchange. (See Blankespoor 2022, Lee and Zhong 2022 and Wong et al. 2024 for more details.) The Cloud Interview column, as part of the EasyIR website, is a specialized channel for real-time text-based communication including OEC.	294 (10.8%)
IR League	**T0/T1:** Allows for customized username only.	Commercial service built on WeChat Mobile App. No PC interface.	81 (3.0%)
Shanghai Securities News	**T0/T1:** Sign up using phone number, and phone number serves as username. No profile customization allowed.	A state-run financial media and CSRC-designated channel for disclosing corporate filings.	75 (2.8%)
China Securities Journal	**T0/T1:** Allows for customized user name only.	A state-run financial media and CSRC-designated channel for disclosing corporate filings.	33 (1.2%)
Straight Flush	**T0/T1+T2:** Allows for both user-uploaded profile image and customized username.	Commercial service provided by Straight Flush, a financial news, social media, data and software firm, and a broker. Straight Flush itself is listed in Shenzhen Stock Exchange (SZ.300033) and typically uses this channel for its own OECs.	28 (1.0%)
Eastmoney	**T0/T1+T2:** Allows for both user-uploaded profile image and customized username.	Commercial service provided by Eastmoney, a financial news, social media, data and software firm, and a broker. Eastmoney itself is listed in Shenzhen Stock Exchange (SZ.300059) and typically uses this channel for its own OECs.	24 (0.9%)

This table outlines the treatment level we exert on each of the nine selected platforms and their notable features. Additionally, we list the total count and corresponding fraction of sample OECs on each platform.

Table 3. Design of Treatments

	T0: Control	T1: Text
Question 1	公司本期盈利水平如何？	【查询-001】高管您好，请问贵公司本期财务报告中，盈利表现如何？谢谢。
	How was the firm's profitability during this reporting period?	[Query-001] Hello, could you please tell me about the firm's profitability in the current financial report? Thank you.
Question 2	公司之后的盈利有什么增长点？	【查询-002】高管您好，请问贵公司未来盈利增长的主要驱动因素有哪些？谢谢。
	What are the potential drivers of the company's future profitability growth?	[Query-002] Hello, could you please share the main drivers of your company's future profit growth? Thank you.
Question 3	你们行业本期整体业绩怎么样？你们跟其他公司比如何？	【查询-003】高管您好，能否请您介绍一下本期行业整体和行业内其他主要企业的业绩表现？谢谢。
	How did your industry perform overall during this reporting period? How does your company compare with your industry peers?	[Query-003] Hello, could you please comment on the overall performance of the industry and other major players during this reporting period? Thank you.
Question 4	行业以后的发展前景怎样？	【查询-004】高管您好，请问您如何看待行业未来的发展前景？谢谢。
	What do you think of the future prospects of the industry?	[Query-004] Hello, what do you think of the future prospects of the industry? Thank you.

T2: Image

AI investors uniformly change their profile picture to the image below and pose questions to the management in a robotic narrative (T1: Text):



Table 4. Summary Statistics

Panel A. Firm (conference)

Variable	N	mean	Std	min	p5	p25	median	p75	p95	max
Treated (Text)	2,722	0.503	0.500	0.000	0.000	0.000	1.000	1.000	1.000	1.000
Image	2,722	0.216	0.412	0.000	0.000	0.000	0.000	0.000	1.000	1.000
Age	2,722	23.744	6.401	9.120	13.394	19.318	23.718	27.855	33.199	40.633
Size	2,709	22.511	1.001	20.199	21.201	21.806	22.321	23.054	24.421	27.021
Book to Market	2,709	0.554	0.382	0.042	0.122	0.300	0.461	0.699	1.289	2.172
Leverage	2,722	0.416	0.212	0.058	0.096	0.246	0.398	0.564	0.793	0.929
Annual Return	2,655	-0.014	0.299	-0.514	-0.391	-0.206	-0.064	0.107	0.534	1.260
Return Volatility	2,696	0.149	0.063	0.060	0.078	0.111	0.136	0.172	0.268	0.448
ROA	2,722	0.014	0.065	-0.253	-0.108	-0.005	0.022	0.051	0.100	0.147
Performance Volatility	2,555	0.040	0.042	0.001	0.004	0.014	0.027	0.052	0.121	0.247
Loss	2,722	0.264	0.441	0.000	0.000	0.000	0.000	1.000	1.000	1.000
ΔROA	2,669	-0.010	0.049	-0.216	-0.091	-0.024	-0.004	0.007	0.058	0.148
EPS Rise	2,655	0.444	0.497	0.000	0.000	0.000	0.000	1.000	1.000	1.000
Abs. ΔEPS	2,655	0.030	0.050	0.000	0.001	0.005	0.013	0.032	0.123	0.308
Earnings Surprise	2,682	-0.016	0.051	-0.277	-0.099	-0.025	-0.008	0.002	0.039	0.132
Analyst Forecast	2,682	0.532	0.871	-1.884	-0.576	0.080	0.380	0.865	2.115	4.103
Analysts Following	2,722	0.991	1.106	0.000	0.000	0.000	0.693	1.946	3.045	4.143
Inst. Ownership	2,718	0.383	0.251	0.003	0.023	0.152	0.379	0.589	0.801	0.893
Board Independence	2,722	0.384	0.054	0.333	0.333	0.333	0.375	0.429	0.500	0.571
R&D Intensity	2,722	0.067	0.085	0.000	0.000	0.020	0.043	0.075	0.227	0.517
SOE	2,722	0.253	0.435	0.000	0.000	0.000	0.000	1.000	1.000	1.000
Customer Concentration	2,663	0.366	0.232	0.021	0.066	0.183	0.320	0.518	0.820	0.960
Supplier Concentration	2,667	0.352	0.192	0.065	0.109	0.204	0.311	0.466	0.745	0.894
# Q&As	2,720	11.172	9.164	0.000	1.000	5.000	9.000	15.000	33.000	40.000
Duration (minutes)	2,720	87.445	27.651	30.000	60.000	60.000	90.000	120.000	120.000	120.000
# Responses	2,722	2.978	1.411	0.000	0.000	2.000	4.000	4.000	4.000	4.000
Non-Response	2,722	0.117	0.321	0.000	0.000	0.000	0.000	0.000	1.000	1.000

Panel B. Firm-question										
Variable	N	mean	Std	min	p5	p25	median	p75	p95	max
Response	10,888	0.744	0.436	0.000	0.000	0.000	1.000	1.000	1.000	1.000
Sentiment	8,106	0.077	0.053	-0.065	-0.018	0.042	0.083	0.115	0.157	0.190
Neutrality	8,106	-0.088	0.040	-0.190	-0.158	-0.116	-0.088	-0.057	-0.025	-0.008
Response Time (minutes)	7,799	34.976	31.528	3.000	6.000	13.000	25.000	45.000	99.000	174.000
# Characters	8,106	183.328	139.170	22.000	38.000	86.000	147.000	237.000	459.000	751.950
# Words	8,106	96.239	75.486	12.000	20.000	44.000	74.000	124.000	247.000	408.950
# Sentences	8,106	3.875	2.467	1.000	1.000	2.000	3.000	5.000	9.000	14.000

This table summarizes our sample. Panel A shows our conference (firm)-level variables, and Panel B shows our experimental firm-questions. For each variable in both panels, we compute the count (N), mean, standard deviation (std), minimum value (min), maximum value (max), and values at the 5th (p5), 25th (p25), 50th (median), 75th (p75), and 95th (p95) percentiles. All continuous and count variables except for Size, Analysts Following, Duration, and # Responses are winsorized at the 1% and 99% levels before calculating their summary statistics. See Table Variable Definitions for detailed definitions of all these variables.

Table 5. Balance Checks

Variable	mean(1)	N(1)	mean(0)	N(0)	mean(1)-mean(0)
Return Volatility	0.153	1,300	0.146	1,288	0.007
Loss	0.298	1,300	0.252	1,288	0.046
Inst. Ownership	0.365	1,300	0.390	1,288	-0.025
Leverage	0.419	1,300	0.403	1,288	0.017
Abs. ΔEPS	0.032	1,300	0.029	1,288	0.004
EPS Rise	0.426	1,300	0.452	1,288	-0.026
Book to Market	0.536	1,300	0.554	1,288	-0.018
Size	22.461	1,300	22.506	1,288	-0.045
Customer Concentration	0.366	1,300	0.356	1,288	0.010
Supplier Concentration	0.354	1,300	0.346	1,288	0.008
Analysts Following	0.981	1,300	1.026	1,288	-0.044
SOE	0.238	1,300	0.255	1,288	-0.017
R&D Intensity	0.069	1,300	0.067	1,288	0.002
Annual Return	0.383	1,300	0.384	1,288	-0.001
Board Independence	-0.016	1,300	-0.021	1,288	0.005

This table provides balance checks for pre-experiment firm (conference)-level covariates used as control variables in our baseline regressions. For each variable, we group observations into the two treatment states and report the count and mean of each group. We first conduct Welch's *t*-test independently for each covariate. Next, we apply the BY procedure (Benjamini and Yekutieli 2001), which downward-adjusts the critical value of statistical significance respectively for each covariate. Rows are sorted according to original *p*-values of differences in means with ascending order. *, **, and *** denote statistical significance at BY-adjusted critical values of 10%, 5%, and 1%, respectively. Table Variable Definitions provides detailed definitions of these variables.

Table 6. Perceived AI Bots and Managers' Likelihood of Responding

Panel A. Full sample				
	(1)	(2)	(3)	(4)
Treated=1	-0.132***		-0.132***	
	(-10.661)		(-10.695)	
Text=1		-0.064***		-0.064***
		(-3.912)		(-3.931)
Text # Image=1		-0.166***		-0.166***
		(-5.800)		(-5.797)
		10,352 observations		
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Batch FE	Y	Y	Y	Y
Platform FE	Y	Y	Y	Y
Question FE			Y	Y

Panel B. By question								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Q1		Q2		Q3		Q4	
Treated=1	-0.132***		-0.094***		-0.149***		-0.153***	
	(-10.253)		(-6.280)		(-8.333)		(-9.277)	
Text=1		-0.075***		-0.027		-0.087***		-0.067***
		(-4.312)		(-1.362)		(-3.729)		(-3.079)
Text # Image=1		-0.136***		-0.162***		-0.147***		-0.215***
		(-4.152)		(-4.555)		(-3.937)		(-5.722)
				2,588 observations				
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y
Batch FE	Y	Y	Y	Y	Y	Y	Y	Y
Platform FE	Y	Y	Y	Y	Y	Y	Y	Y

In this table, we estimate the treatment effects on managers' likelihood of responding using Probit regressions. Treated and Text equal 1 for treated firm-questions, and 0 otherwise. Image is 1 for questions with an explicit AI bot label, and 0 otherwise. Control variables are Size, Book to Market, Leverage, Annual Return, Return Volatility, Loss, EPS Rise, Abs. ΔEPS, Analysts Following, Inst. Ownership, Board Independence, R&D Intensity, SOE, Customer Concentration and Supplier Concentration. Table Variable Definitions provides detailed variable definitions. In Panel A, we pool all firm-question observations and cluster standard errors by firm, and in Panel B, we regress respectively for each of our four questions using robust standard errors. This table reports average marginal effects. *t*-statistics are in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 7. Heterogeneity in Treatment Effects

	(1)	(2)	(3)	(4)
Indicator:	SOE	Female	Elderly	Computer Science
Indicator # Treated=1	0.072***	-0.132*	-0.105**	-0.469***
	(2.939)	(-1.669)	(-2.115)	(-6.943)
Indicator=1	-0.026	0.091***	0.021	0.154***
	(-1.100)	(2.793)	(0.751)	(3.578)
Treated=1	-0.150***	-0.051***	-0.036*	-0.024
	(-10.610)	(-2.813)	(-1.742)	(-0.523)
Observations	10,352	4,236	4,236	740
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Batch FE	Y	Y	Y	Y
Platform FE	Y	Y	Y	Y
Question FE	Y	Y	Y	Y

In this table, we examine cross-sections of treatment effects on managers' likelihood of responding using Probit regressions with Treated interacted with an indicator variable. The indicators are SOE, Female, Elderly, and Computer Science in columns (1), (2), (3), and (4), respectively. Treated equals 1 for treated firm-questions, and 0 otherwise. Elderly equals 1 if the responding manager's age is among the top 25% in the analysis sample, and 0 otherwise. Control variables are Size, Book to Market, Leverage, Annual Return, Return Volatility, Loss, EPS Rise, Abs. ΔEPS, Analysts Following, Inst. Ownership, Board Independence, R&D Intensity, SOE, Customer Concentration, and Supplier Concentration. Table Variable Definitions provides detailed variable definitions. In column (1), the analysis sample is all firm-questions with non-missing controls. In columns (2) to (4), the sample is restricted to OECs on SSE Roadshow and Shanghai Securities News, where we post our questions with an explicitly designated manager as the respondent. This table reports average marginal effects. *t*-statistics are in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are clustered by firm.

Table 8. Perceived AI Bots and Managerial Sentiment Manipulation

	(1) Sentiment	(2) Neutrality
Treated=1	-0.007***	0.005***
	(-3.737)	(3.411)
IMR	0.023***	-0.011**
	(3.260)	(-1.987)
	7,417 observations	
Controls	Y	Y
Industry FE	Y	Y
Batch FE	Y	Y
Platform FE	Y	Y
Question FE	Y	Y

This table reports the results of the second-step regressions of Heckman (1976)'s approach. The outcome variable is Sentiment in columns (1) and (2) and Neutrality in columns (3) and (4). Treated equals 1 for treated firm-questions, and 0 otherwise. Control variables consist of sentiment predictors from Huang et al. (2014): Age, Size, Book to Market, Annual Return, Return Volatility, ROA, Performance Volatility, Loss, ΔROA, Earnings Surprise, and Analyst Forecast. Table Variable Definitions provides detailed variable definitions. IMR is the inverse Mills ratio derived from Equation (5). *t*-statistics are in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are clustered at the firm level.

Table 9. Do Managers Take AI-Styled Questions More Seriously?

	(1) log. # Characters	(2) log. # Words	(3) log. # Sentences	(4) Response Time
Treated=1	0.093***	0.089***	0.063**	-2.870*
	(2.638)	(2.607)	(2.183)	(-1.702)
IMR	-0.050	-0.006	0.024	9.282
	(-0.352)	(-0.044)	(0.207)	(1.272)
Observations	7,704	7,704	7,704	7,434
Industry FE	Y	Y	Y	Y
Batch FE	Y	Y	Y	Y
Platform FE	Y	Y	Y	Y
Question FE	Y	Y	Y	Y

This table reports the results of the second-step regressions of Heckman (1976)'s approach. The outcome variable is the natural logarithm of # Characters, # Words, and # Sentences in column (1), (2), and (3), respectively. In column (4), the outcome variable is Response Time. Treated equals 1 for treated firm-questions, and 0 otherwise. Table Variable Definitions provides detailed variable definitions. IMR is the corresponding inverse Mills ratio derived from Equation (5). *t*-statistics are in parentheses. *, **, and *** denote statistical significance at 10%, 5%, and 1% level, respectively. Standard errors are clustered by firm.

Table 10. Market Responses

	(1)	(2)	(3)	(4)
	Question arrival		Managerial response	
	[-1min, 1min]	[-2min, 2min]	[-1min, 1min]	[-2min, 2min]
Treated=1	-3.609*	-4.793*	0.445	-0.569
	(-1.703)	(-1.804)	(0.472)	(-0.451)
Observations	1,329		1,667	
Contemp. shocks controlled	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Batch FE	Y	Y	Y	Y
Question FE	Y	Y	Y	Y

This table examines the market response to our experimental interventions using only the SSE Roadshow sample. The dependent variable is AbnVolume (see Table Variable Definitions for detailed construction). We construct the examined window around the arrival of our experimental question in columns (1) and (2) and around the timestamp of the managerial response in columns (3) and (4). We control for contemporaneous shocks from real-world questions from other OEC participants. The window spans are [-1min, 1min] in columns (1) and (3) and [-2min, 2min] in columns (2) and (4). Only firm-questions during trading hours (9:30 to 11:30 and 13:00 to 14:56) are included. *, **, and *** denote statistical significance at 10%, 5%, and 1% levels, respectively. Standard errors are clustered by firm.

Online Appendix A. Validations of Treatments

In the RCT, we raised questions to corporate managers on 2,722 online earnings conferences (OECs) hosted by listed firms in China. We exerted three treatment layers: human-style question prompts simulating average OEC participants (T0), undisguised and AI-styled question prompts nudging managers to perceive presence of AI-readership (T1), and a robotic icon where profile image is modifiable in addition to AI-styled questions (T1+T2).

We are going to check our RCT manipulation (T0, T1, and T1+T2) via a lab experiment, trying to examine what otherwise remains unexplored in the RCT setting due to institutional constraints. Specifically, we aim to see if (1) T1 alone is sufficient to trigger perceptions of AI-readership; (2) T2 alone (i.e., a robotic icon) in absence of T1 can have an effect; and (3) each of the treatment arms (T0, T1, and T1+T2) shapes perceived intelligence and trustworthiness of questioners.

Transparency

This study was approved by the School of Business, Renmin University of China (protocol #2026R24). The following URL provides a complete demonstration of the experiment procedure: https://www.credamo.com/s/rqIFVfano.

Participants

A total of 240 participants recruited from Credamo (https://www.credamo.com/#/) completed this study online in May 2026 in exchange for a small monetary reward. Note that we were neither able to perfectly replicate an OEC in the lab nor to recruit a sample of firm managers, in this study, we posed no demographic restriction on participants, providing suggestive evidence on how the average people responded to our experimental stimulus. We assume that firm managers do not differ systematically from the average people in this setting.

Among the 240 participants, 96 (40.0%) were male, and 2 did not disclose his or her gender. The sample average age was 32.13 years old ($Std = 9.61$). 156 (65.0%) held a

bachelor's degree, 51 (21.25%) a master's degree, and 3 (1.25%) a doctoral degree. 109 (45.42%) indicated that they were experienced in security trading, with an average of 5 ($Std = 1.83$) frequently used social media platforms and an average of 2 ($Std = 1.33$) frequently used commercial AI tools. All participants passed the attention check.

Procedure and experiment conditions

After providing informed consent, each participant saw a mockup screenshot of comments below a post from Guba (https://guba.eastmoney.com/), a leading and dedicated social media platform for financial investment in China. We adopt the Guba setting because (1) a noticeable proportion of Guba users use default usernames and profile image, a feature also seen in the OEC setting and partly constituting our RCT manipulation, (2) Guba features a userbase with certain level of financial sophistication similar to the average OEC participants, and (3) participants are more familiar with the Guba setting instead of the OEC setting. We deceptively told the participants that they were viewing an authentic webpage screenshot in order to increase apparent realism of this study.

We employed a 2×2 between-subject design ($N = 4 \times 60$) with linguistic style and profile image as the two manipulated independent variables, yielding a total of four experiment conditions. Each participant received one of the four conditions with identical probability. The mockup screenshot remained almost the same across all four conditions, except for the two pieces of comments from one specific user and her profile image. This manipulated user's comments were linguistically human-like under conditions T0 and T0+T1 and linguistically robotic (including exactly the same "[Query-00*x*]" prefix as in the RCT and *x* can be either 1 or 2) under conditions T1 and T1+T2; she used the default profile image under conditions T0 and T1 and a robotic icon (exactly the same as in the RCT) under conditions T0+T2 and T1+T2. For consistency and ease of understanding, we adopt the same T0, T1, and T2 notation as in the RCT; however, one should keep in mind that in this lab environment, we were able to exert

the T2 treatment in absence of T1, which was fundamentally different from the RCT setting.

After viewing the mockup screenshot, all participants were first told to mark the username of any Guba user they felt weird or different from others by clicking. Next, we explicitly asked the participants to click on the username of any user they suspect of a robot, algorithm, or AI agent identity. Then, we directed participants' attention to the manipulated user, and asked them about the user's username, which served as an attention check. Finally, the participants filled out a scale measuring their perceptions of the user's intelligence and trustworthiness.

Measures

Perceived weirdness and robotic identity

In the RCT, we claimed that robotic linguistic style and the prefix and a robotic profile image could trigger perceived identity of robot, algorithm, or AI agent. If this was true, we should observe that the manipulated user received more clicks under conditions T0+T2, T1, and T1+T2, compared to the benchmark T0 condition. We code the measure as 1 if a participant clicked on the manipulated user and 0 otherwise.

Besides, we included a nudging question on perceived weird user(s) before formally and explicitly asking participants about perceived user(s) with a robotic identity. Again, we code the measure as 1 if a participant clicked on the manipulated user and 0 otherwise. We included this question because perception of robotic identity could be triggered simply by our question prompt. We expect participants to respond similarly to these two questions, since the manipulated user's robotic traits could contribute to a feeling of weirdness.

Perceived intelligence and trustworthiness

We measure participants' perceptions of the manipulated user's intelligence and trustworthiness using a 11-item and 6-point ($1 = strongly\ disagree$ and $6 = strongly\ agree$) scale (Cronbach's $\alpha = 0.947$) adapted from McCroskey and Teven's (1999) measure of source credibility. We considered items belonging only to the intelligence

(Cronbach's $\alpha = 0.890$) or the trustworthiness (Cronbach's $\alpha = 0.919$) dimension. We take the simple averages of item scores to compute our intelligence, trustworthiness, and source credibility measures.

As suggested, we employed a three-round back-translation process. (Brislin 1970; Klotz et al. 2023) Specifically, since the original scale items comprised only of adjective pairs (McCroskey and Teven 1999), we retained only the lefthand-side adjectives, dropped redundant synonyms, and had one of the authors translate remaining adjectives into 11 Chinese words. Next, each translated word was inserted into a Chinese declarative sentence depicting the manipulated user. Then, we directly prompted ChatGPT to translate the 11 Chinese sentences back into English in a temporary chat, leveraging the feature that model memories were not shared among different temporary chats. Finally, we compared the adjectives in the translated sentences to the original scale items and made reasonable adaptions to our Chinese translation. We repeated this procedure three times. A confirmatory factor analysis finds that the two-factor model has a satisfactory model fit ($CFI = 0.983,\ RMSEA = 0.056,\ SRMR = 0.027$).

Results

Checks of the RCT manipulation

We begin by looking at the spatial distribution of participant clicks when answering our question about perceived weird users on the mockup Guba webpage. In Figure OA1, we draw a heatmap of clicks from participants under each of our four experiment conditions. We crop the full heatmaps for brevity here, since the original figures shown to the participants are too large in size. The comments from the manipulated user are highlighted by blue rectangles. Compared panel-wise, we observe that the manipulated user received much fewer clicks under condition T0 compared to all the other three conditions, suggesting that this user seemed much weirder if she used either robotically phrased language, a robotic icon as her profile image, or

both. Figure OA2 documents the spatial distribution of participant clicks on perceived robotic users and reveals a similar pattern: participants clicked more on the manipulated user under conditions T1, T0+T2, and T1+T2, relative to the benchmark T0 condition. Taken together, both robotic language usage and a robotic icon as the profile image can lead to more clicks from the participants, and this pattern is consistent no matter we explicitly mentioned robotic identity or not.

Next, we formally test for our observations from Figure OA1 and Figure OA2 using the following Probit regressions:

$$Perception_i = \alpha + \gamma \times Text_i + \delta \times Image_i + \zeta \times Text_i \times Image_i + \varepsilon_i, \qquad (a1)$$

where i denotes participant, and $Perception_i$ is a dummy variable equal to 1 if the participant clicked on the manipulated Guba user when answering either the perceived weird users or the perceived robotic user question. $Text_i$ and $Image_i$ are dummy variables and indicate the experiment condition a participant received: in absence of the other, $Text_i$ ($Image_i$) equals 1 if the participant received the T1 or T1+T2 (T0+T2 or T1+T2) condition; and when joining the equation together, $Text_i$ ($Image_i$) equals 1 if the participant received the T1 (T0+T2) condition, and the ostensible interaction term, $Text_i \times Image_i$, equals 1 if the participant received the T1+T2 condition. Heteroscedasticity-robust standard errors are used here.

Table OA1 reports the results. In support of our conjecture, all coefficients are statistically distinguished from 0 with large magnitudes. Columns (1) to (3) indicate an increase by more than 20 percentage points in the probability that the manipulated user was perceived weird when she used either robotic language or a robotic profile image, and when she used both the two instruments, the increase is greater than 40%. Columns (4) to (6) focus on the perceived robotic identity of the manipulated user and document a similar pattern. However, although statistically significant, the impact of a robotic profile image is much smaller in magnitude,

around only 13% compared to the more than 40% increase generated by robotic language. This suggests that robotic language, even in absence of a visual hint, is a primary contributor of robotic perception. Overall, our results are consistent with a successful manipulation in our RCT.

Perceived intelligence and trustworthiness

We also examine whether and how our treatment arms shaped perceived intelligence and trustworthiness of the manipulated Guba user. Specifically, we run OLS regressions with all other specifications remained the same as in equation (a1) except that the outcome variable is perceived intelligence, trustworthiness, or source credibility. Results are reported in Table OA2. In column (2) of Table OA2, we observe a 0.40 (9.67% relative to the sample mean of 4.13) increase in perceived intelligence when the manipulated user used a robotic profile image. As is evident in column (3), the impact of this image is primarily driven by the T1+T2 condition, where the linguistic dimension was also included as a trigger of robotic perception. Columns (7) to (9) of Table OA2 reveal a similar pattern for perceived source credibility, with a magnitude of 0.33 (8.03%) relative to the sample mean of 4.06. As expected, through columns (4) to (6), we observe little, and even positive variations in perceived trustworthiness resulting from our T2 treatment, contradictory to the notion that our observed decline in response propensity in the RCT is primarily due to managers' interpretation of AI-styled questioner as untrustworthy. Moreover, we find that robotic linguistic alone is never sufficient to yield meaningful variations in perceived intelligence, trustworthiness, or source credibility. By varying linguistic in our RCT, we inevitably introduce variations in other dimensions such as if the questioner is regarded an institutional investor or not. We alleviate this concern in that, at least in our lab setting, people did not form such belief based on simple linguistic hints.

Figure OA1. Participant Clicks on Perceived Weird Users

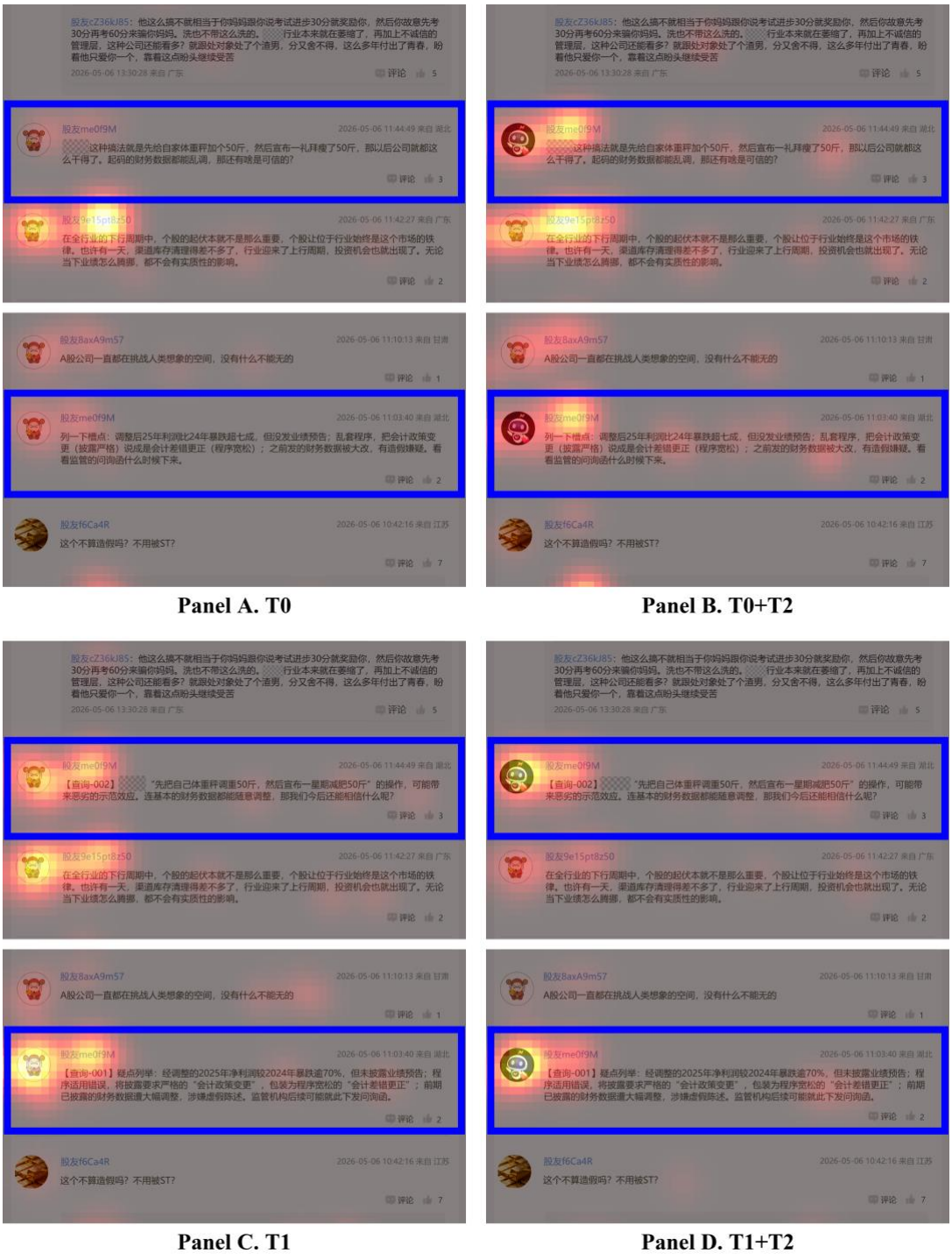


Panel A. T0 **Panel B. T0+T2**

Panel C. T1 **Panel D. T1+T2**

This figure shows a heatmap of clicks from participants under each of our four experiment conditions. Panels A, B, C, and D are for conditions T0, T0+T1, T1, and T1+T2, respectively. Each heatmap shown here is cropped and displays comments only from the manipulated Guba user or chronologically beside. The manipulated user's comments are highlighted by blue rectangles. Lighter (yellower) colors, relative to darker (redder) ones, denote more intensive clicks from participants. Other shaded regions, neither yellow, orange, nor red, receive no click.

Figure OA2. Participant Clicks on Perceived Robotic Users



Panel A. T0 **Panel B. T0+T2**

Panel C. T1 **Panel D. T1+T2**

This figure shows a heatmap of clicks from participants under each of our four experiment conditions. Panels A, B, C, and D are for conditions T0, T0+T1, T1, and T1+T2, respectively. Each heatmap shown here is cropped and displays comments only from the manipulated Guba user or chronologically beside. The manipulated user's comments are highlighted by blue rectangles. Lighter (yellower) colors, relative to darker (redder) ones, denote more intensive clicks from participants. Other shaded regions, neither yellow, orange, nor red, receive no click.

Table OA1. Formal Checks of the RCT Manipulation

	(1)	(2)	(3)	(4)	(5)	(6)
Manipulated user perceived:		Weird			Robotic	
Text=1	0.200***		0.211***	0.467***		0.403***
	[120]		[60]	[120]		[60]
Image=1		0.233***	0.241***		0.133**	0.141**
		[120]	[60]		[120]	[60]
Text # Image=1			0.413***			0.477***
			[60]			[60]
			240 observations			

In this table, we run Probit regressions of a dummy on experiment condition indicators. The outcome dummy is whether a participant clicked on the manipulated Guba user in the nudging question (i.e., the participant perceived this user weird) in columns (1) to (3) and in the question explicitly asking about robotic identity (i.e., the participant perceived this user robotic) in columns (4) to (6). In columns (1) and (4) ((2) and (5)), Text (Image) equals 1 if the participant received either the T1 or T1+T2 (T0+T2 or T1+T2) condition, and 0 otherwise. In columns (3) and (6), Text equals 1 if the participant received the T1 condition and 0 otherwise, and Image equals 1 if the participant received the T0+T2 condition and 0 otherwise. The ostensible interaction term in columns (3) and (6), Text # Image, equals 1 if the participant received the T1+T2 condition, and 0 otherwise. This table reports average marginal effects. For clarity, we place the number of participants receiving the experiment condition(s) denoted by either Text, Image, or Text # Image in parenthesis. *, **, and *** denote statistical significance at 10%, 5%, and 1% level. Robust standard errors are used.

Table OA2. Perceived Intelligence and Trustworthiness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Manipulated user perceived:		Intelligent			Trustworthy			Credible	
Text=1	0.013		-0.167	-0.033		-0.222	-0.012		-0.197
	[120]		[60]	[120]		[60]	[120]		[60]
Image=1		0.400***	0.220		0.286**	0.097		0.338***	0.153
		[120]	[60]		[120]	[60]		[120]	[60]
Text # Image=1			0.413**			0.253			0.326*
			[60]			[60]			[60]
240 observations									

In this table, we regress the manipulated Guba user's perceived traits by the participant on experiment condition indicators. The trait we consider is intelligence in columns (1) to (3), trustworthiness in columns (4) to (6), and source credibility, measured via an adaption of McCroskey and Teven's (1999) scale of source credibility and taking the simple average of item scores. In columns (1), (4) and (7) ((2), (5) and (8)), Text (Image) equals 1 if the participant received either the T1 or T1+T2 (T0+T2 or T1+T2) condition, and 0 otherwise. In columns (3), (6) and (9), Text equals 1 if the participant received the T1 condition and 0 otherwise, and Image equals 1 if the participant received the T0+T2 condition and 0 otherwise. The ostensible interaction term in columns (3), (6) and (9), Text # Image, equals 1 if the participant received the T1+T2 condition, and 0 otherwise. For clarity, we place the number of participants receiving the experiment condition(s) denoted by either Text, Image, or Text # Image in parenthesis. *, **, and *** denote statistical significance at 10%, 5%, and 1% level. Robust standard errors are used.

Online Appendix B. Alternative Validation of Treatments via LLMs

1. The Original Interactions with DeepSeek-R1 (in Chinese)

Training Section:

用户：

下面，我将给你四对问题，每一对问题均涵盖相同的内容，但采用不同的表达方式。请为每一对中的两个问题分别打分：1 分（几乎确定是人写的）；2 分（可能是人写的）；3 分（可能是机器生成的）；4（几乎确定是机器生成的）。为每一个问题分别给出你的打分理由。

DeepSeek-R1：

已经深度思考（用时 61 秒）

好的，我现在需要处理用户的请求，给四对问题中的每个问题打分，并给出理由。首先，我得仔细理解用户的需求。用户提供的问题对都是相同的内容，但表达方式不同。我的任务是为每个问题分别打分，1 分是人写的，4 分是机器生成的。同时，每个评分都需要理由。

首先，我需要明确评分标准。1 分是几乎确定是人写的，可能更自然，有情感或个性化表达。2 分可能是人写的，但可能有点生硬或不流畅。3 分可能是机器生成的，结构正确但缺乏自然感。4 分几乎确定是机器生成的，可能有重复、不自然的结构或用词不当。

接下来，我要分析每一对问题。比如，第一对可能涉及不同表达方式，比如一个更口语化，另一个更正式。需要检查是否有重复的短语，结构是否机械，用词是否准确自然。

然后，我需要逐对分析。比如，第一个问题对，可能一个是直接问“如何学习英语”，另一个可能用更复杂的结构，比如“有哪些方法可以有效地学习英语”。这时候要判断哪个更自然，哪个可能像机器生成的。

同时，要注意中文表达的特点。比如，口语中常用缩略语或省略主语，而机器生成的可能更完整但生硬。比如，使用“建议”、“首先”等结构可能更机械，而人类可能会用更随意的表达。

另外，要注意是否有语法错误或用词不当。机器生成的文本可能会有不恰当的词序或用词，而人类写的会更流畅。但有时候，人类也可能写出结构较生硬的句子，所以需要综合考虑。

对于每个问题，我需要分别打分，并给出理由。比如，如果一个问题使用了模板化的结构，重复使用相同的句式，或者用词不自然，可能倾向于 3 或 4 分。而如果问题有个人经验或情感，更可能是 1 或 2 分。

还要注意用户可能希望评分细致，所以每个理由需要具体指出问题所在，比如指出用词、结构、流畅度等方面的问题。

现在，我需要模拟处理四对问题，虽然用户没有给出具体内容，但根据常见情况，假设每对中的问题有不同特点。例如：

第一对：

问题 A：如何快速学习英语？

问题 B：有哪些有效的方法可以提升英语学习效率？

可能问题 A 更简洁，直接，像人类提问；问题 B 结构更复杂，可能机器生成，但不确定，需要分析用词和结构。

第二对：

问题 C：学英语有什么好方法吗？

问题 D：请提供一些高效的英语学习方法建议。

问题 C 更口语化，可能 1 分；问题 D 更正式，可能机器生成，得 3 或 4 分。

第三对：

问题 E：怎样才能提高英语水平？

问题 F：为了提升英语能力，应采取哪些学习策略？

问题 E 自然，可能 1 分；问题 F 用词正式，结构严谨，可能机器生成，得 3 分。

第四对：

问题 G：有什么技巧可以学好英语？

问题 H：请列举几种能够有效促进英语学习的技巧。

问题 G 自然，得 1 分；问题 H 重复结构，用词正式，可能得 4 分。

不过这只是假设，实际需要根据具体问题分析。但用户没有给出具体问题，可能需要用户提供后评分。不过用户可能希望我以示例形式展示处理过程，但根据指示，用户已经给出任务，但问题对未列出，可能用户输入有误，或者需要我假设。但根据原始查询，用户可能是在测试我的评分能力，所以需要明确步骤，然后示例处理。

但根据用户的问题描述，用户确实会提供四对问题，但当前输入中可能没有，可能用户输入不完整。不过根据现有信息，可能用户希望我展示如何处理，即使没有具体问题。但根据指示，可能用户希望我等待他提供问题对，但原帖中可能遗漏了。需要检查用户的历史输入。

查看用户的历史消息，用户确实在问题中提到了四对问题，但实际内容可能未正确粘贴，导致问题对缺失。在这种情况下，可能需要用户补充问题对，但作为助手，我需要处理现有信息。可能用户希望我提供一个示例回答，或者指出问题对缺失。但根据指示，用户可能希望我继续处理，假设问题对存在，并给出评分示例。但这样可能不准确。

或者，可能用户的问题存在格式问题，实际的问题对被包含在后续内容中，但未被正确解析。需要仔细检查原始输入。

原始用户输入中，问题描述后是否有问题对？根据用户的消息：“下面，我将给你四对问题，每一对问题均涵盖相同的内容，但采用不同的表达方式。请为每一对中的两个问题分别打分...”但之后没有列出具体的问题对。这可能是一个错误，用户可能忘记粘贴问题对，或者格式问题导致未显示。

在这种情况下，作为助手，我需要指出问题对缺失，无法进行评分。但用户可能希望我继续，假设问题对存在，并提供一个示例回答。或者，可能用户的问题对在后续的输入中，但被系统分割。需要确认。

由于当前信息不完整，无法进行准确的评分。但根据问题描述，用户可能希望我展示如何处理这类任务。因此，可能需要说明如果没有具体问题对，无法进行评分，并请求用户提供。但根据指示，用户可能希望我继续，因此可能需要构造一个假设的例子来说明评分过程。

例如，假设用户提供了以下四对问题：

第一对：

问题 1A：怎么才能快速学好英语？

问题 1B：有哪些方法可以有效提升英语学习效率？

第二对：

问题 2A：学英语有什么好方法吗？

问题 2B：请提供一些高效的英语学习方法建议。

第三对：

问题 3A：怎样提高英语听力？

问题 3B：如何有效地提升英语听力能力？

第四对：

问题 4A：有什么好的英语学习技巧？

问题 4B：请列举几种能够有效促进英语学习的技巧。

针对这些假设的问题对，进行评分：

对于问题 1A：“怎么才能快速学好英语？”这个句子结构简单，口语化，符合人类自然提问方式，可能得 1 分。

问题 1B：“有哪些方法可以有效提升英语学习效率？”结构稍正式，但常见于人类写作，可能得 2 分，或 3 分，若结构机械。
但需要更详细分析。例如，“有效提升...效率”可能更正式，但并非机器生成的特征，可能得 2 分。
评分理由需要具体，比如问题 1A 使用口语化的“怎么才能”和“学好”，而问题 1B 使用“有效提升...效率”，结构完整，可能更书面，但并非明显机器生成。
但机器生成的问题可能使用重复结构或特定模板，比如“有哪些方法可以...”，这可能常见于人类和机器生成，需要更多线索。
可能更准确的是，问题 1A 得 1 分，理由：自然口语化，结构简单；问题 1B 得 2 分，结构稍正式但合理。
假设用户的问题对类似这样的情况，我需要为每对中的每个问题评分并给出理由。
总结来说，处理步骤是：
1. 分析每个问题的语言特征，如口语化程度、结构复杂度、用词选择。
2. 对比常见的人类和机器生成文本的特点。
3. 根据评分标准，确定每个问题的得分。
4. 撰写每个评分的理由，指出具体的用词、结构或流畅度问题。
需要注意避免偏见，比如较长的句子不一定是机器生成，但结构重复或用词不自然可能是机器特征。同时，中文的机器生成文本可能有特定的模式，如过度使用某些短语或结构。
最终，用户可能需要一个结构清晰的回答，对每对问题分别列出两个问题的评分和理由，用数字和项目符号排列。
由于您提供的实际“问题对”内容缺失，目前无法进行具体分析和评分。若您能补充具体的四对问题文本，我将按照以下标准为您进行详细评估：
评分标准
1 分（几乎确定是人写的）：口语化表达、自然流畅、存在情感或个性化措辞
2 分（可能是人写的）：结构合理但稍显生硬，偶有不自然词汇
3 分（可能是机器生成的）：语法正确但缺乏自然感，重复句式或模板化结构
4 分（几乎确定是机器生成的）：逻辑混乱、用词不当、明显机械重复或非人性化表达
示例分析（假设问题对）
问题对 1
A：怎么快速学好英语？
评分：1 分
理由：口语化短句，符合人类直接提问习惯，无冗余词汇。
B：有哪些方法可以有效提升英语学习效率？
评分：3 分
理由：结构完整但稍显模板化，“有效提升...效率”属常见机器生成句式。
问题对 2
A：学英语有啥好方法吗？
评分：1 分
理由：使用“有啥”等口语缩略词，自然体现 colloquial 特征。
B：请提供一些高效的英语学习方法建议。
评分：4 分
理由：指令式句型（“请提供”）和正式措辞（“建议”）明显符合机器生成模式。
问题对 3
A：怎样提高英语听力？
评分：2 分
理由：简洁直接，但“怎样”稍显书面化，可能为人工刻意简化。
B：如何有效地提升英语听力能力？
评分：3 分
理由：“有效地...能力”为高频机器搭配，结构正确但冗余。
问题对 4
A：有什么好的英语学习技巧？
评分：1 分
理由：自然提问，使用“好的”等主观评价词，体现人类偏好。
B：请列举几种能够有效促进英语学习的技巧。
评分：4 分
理由：机械性动词（“列举”“促进”）叠加，句式僵硬，属典型机器生成指令。
请提供实际需评分的四对问题文本，我将按此格式进行完整分析。

Testing Section:

用户：
1A：公司本期盈利如何？
1B：【查询-001】高管您好，请问贵公司本期财务报告中，盈利表现如何？
DeepSeek-R1：
已深度思考（用时 20 秒）
嗯，用户这次给了一对问题让我评分，我需要仔细分析每一个的特点。首先看 1A：“公司本期盈利如何？”这句话很简短，直接用了“如何”来提问，非常口语化，像是日常对话中的提问方式。没有复杂的结构，也没有使用任何模板化的表达，感觉像是人类自然问出来的问题。
接下来是 1B：“【查询-001】高管您好，请问贵公司本期财务报告中，盈利表现如何？”这里有几个需要注意的地方。开头有编号“查询-001”，这在日常交流中不太常见，可能是系统生成的编号。称呼“高管您好”显得比较正式，加上“贵公司”这种尊称，通常人类在非正式场合可能不会这么正式，尤其是在内部沟通时。另外，“盈利表现如何”虽然语法正确，但稍微有点冗长，可能更倾向于正式或模板化的表达，比如在官方报告或自动生成的查询中常见。
根据评分标准，1A 的口语化和自然流畅应该得 1 分。而 1B 的结构虽然合理，但编号和正式用语显得生硬，可能是机器生成的，所以评分 2 分或者 3 分？不过用户之前的例子中，类似使用“请提供”和正式措辞的情况得了 4 分，但这里的 1B 虽然有编号和尊称，但整体结构还是通顺的，没有明显的逻辑混乱或用词不当，所以可能更适合 3 分。不过之前的例子中，指令式句型如“请提供”得了 4 分，但这里没有指令动词，只是更正式，所以可能需要再考虑。可能需要比较之前的示例，比如问题对 2 中的 B 用了“请提供”得了 4 分，而这里的 1B 虽然有编号和尊称，但整体结构更接近可能的人类书面沟通，比如正式邮件中的提问，所以可能评 2 分？或者 3 分？
可能需要再仔细对比评分标准。1B 的结构合理但稍显生硬，编号和正式称呼可能让句子显得不够自然，所以可能属于 2 分或者 3 分。根据用户之前的示例，问题对 2 中的 B 因为使用正式措辞和指令式句型得了 4 分，而问题对 3 中的 B 因为高频机器搭配得了 3 分。这里的 1B 没有指令式句型，但用词更正式，且有编号，可能属于 3 分，即可能是机器生成的，因为人类在提问时通常不会加入编号，除非是系统自动生成的查询。因此，可能 1B 评 3 分，而 1A 评 1 分。
以下是对您提供的第一个问题对的评分及理由分析：
问题对 1

1A：公司本期盈利如何？
评分：1 分
理由：
口语化表达，符合日常对话习惯（如直接使用“如何”提问）。
句式简洁自然，无冗余词汇或模板化结构，体现人类提问的直接性和随意性。

1B：【查询-001】高管您好，请问贵公司本期财务报告中，盈利表现如何？
评分：3 分
理由：
存在模板化特征，如编号（【查询-001】）、尊称（“高管您好”“贵公司”）和正式措辞（“财务报告中”“盈利表现”），符合机器生成文本的格式化特征。
语法正确但稍显冗余，如“盈利表现如何”相比 1A 更书面化，缺乏自然对话的简洁性。
类似客服或自动化系统的查询模板，虽逻辑清晰，但缺乏个性化或情感表达。
总结：
1A 符合人类自然提问习惯，而 1B 因格式化编号、尊称叠加及正式措辞，更接近机器生成的标准化查询模板。

2. Translation of Prompt and Response by DeepSeek-R1 (Testing Section Only)

“Below, I will give you four pairs of questions. Each pair covers the same content but is phrased differently. Please score each of the two questions in every pair using the following scale:
1 point – Almost certainly written by a human
2 points – Possibly written by a human
3 points – Possibly machine-generated
4 points – Almost certainly machine-generated
For each question, please also provide a brief justification for your score.

1A: How is the company's profit this period?
1B: [Query-001] Hello Executive, may I ask how your company's profit performance is in the current financial report?”

DeepSeek-R1 (Response):
Thinking deeply (took 20 seconds)

Hmm, this time the user has given me a pair of questions to score, so I need to carefully analyze the characteristics of each. Let's start with 1A:"How is the company's profit this period?" This sentence is very short and uses "how is" to ask a question, which is quite conversational—like something someone would say in everyday dialogue. It doesn't have a complex structure and doesn't use any templated phrasing. It feels like a naturally phrased human question.

Now let's look at 1B: "[Query-001] Hello Executive, may I ask how your company's profit performance is in the current financial report?" There are several things to note here. First, the prefix "Query-001" is uncommon in everyday communication—it looks like a system-generated ID. The salutation “Hello Executive” sounds overly formal, and the phrase “your company” is a polite form that people don't usually use in casual settings, especially in internal communication. Also, “profit performance” is grammatically correct, but it feels slightly redundant or overly formal—more typical of official reports or auto-generated queries.

Final Scores for Pair 1:

1A: How is the company's profit this period?
Score: 1 point
Reason: Informal, conversational phrasing typical of human dialog. Short, direct, and free of templated language. Reflects natural, spontaneous query.

1B: [Query-001] Hello Executive, may I ask how your company's profit performance is in the current financial report?
Score: 3 points
Reason: Features templated elements such as a query number ("[Query-001]"), formal salutation ("Hello Executive"), and overly polished terms ("your company," "profit performance"). Though grammatically correct and logically coherent, it lacks the casual, personal tone of a human speaker. Resembles an automated system-generated query.

Summary:
1A aligns well with natural human questioning habits.
1B, due to its formatting, formal tone, and structured phrasing, more closely resembles a standardized query generated by a machine.

Online Appendix C. Building Blocks for Sentiment Axes

	Positive pole $\mathcal{P}^+$	Negative pole $\mathcal{P}^-$
Question 1	公司本期财务报告中，盈利表现优异。	公司本期财务报告中，盈利表现不佳。
	公司本期财务报告中，盈利表现超出预期。	公司本期财务报告中，盈利表现不达预期。
	公司本期盈利水平超出预期。	公司本期盈利水平不达预期。
	公司本期财务表现优异。	公司本期财务表现不佳。
	公司本期财务表现超出预期。	公司本期财务表现不达预期。
	公司本期业绩非常好。	公司本期业绩不佳。
	公司本期业绩超出预期。	公司本期业绩不达预期。
Question 2	公司未来盈利增长的驱动因素非常丰富。	公司未来盈利增长的驱动因素非常稀缺。
	公司之后的盈利增长点非常丰富。	公司之后的盈利增长点非常稀缺。
	公司未来盈利增长的前景非常广阔。	公司未来盈利增长的前景黯淡。
	公司未来的发展前景非常广阔。	公司未来的发展前景黯淡。
	我们对公司未来盈利增长的前景持乐观态度。	我们对公司未来盈利增长的前景持悲观态度。
	管理层对公司未来盈利增长的前景持乐观态度。	管理层对公司未来盈利增长的前景持悲观态度。
	我们对公司未来的发展前景持乐观态度。	我们对公司未来的发展前景持悲观态度。
	管理层对公司未来的发展前景持乐观态度。	管理层对公司未来的发展前景持悲观态度。
Question 3	公司本期业绩表现领先于全行业。	公司本期业绩表现落后于全行业。
	公司本期业绩表现领跑全行业。	公司本期业绩表现为全行业倒数。
	公司本期业绩表现优于行业平均水平。	公司本期业绩表现比行业平均水平差。
	公司本期业绩相较于行业内其他公司而言处于领先地位。	公司本期业绩相较于行业内其他公司而言处于落后地位。
	公司本期业绩相较于行业内其他公司而言处于领跑地位。	公司本期业绩相较于行业内其他公司而言处于追赶地位。
	公司本期业绩相较行业内其他公司而言非常优秀。	公司本期业绩相较行业内其他公司而言非常差劲。
Question 4	行业未来的发展前景非常广阔。	行业未来的发展前景黯淡。
	行业正在步入上行周期。	行业正在走下坡路。
	行业正在快速发展。	行业正在快速衰落。
	我们对行业未来的发展前景持乐观态度。	我们对行业未来的发展前景持悲观态度。
	管理层对行业未来的发展前景持乐观态度。	管理层对行业未来的发展前景持悲观态度。

Online Appendix D. Additional Empirical Results

Table OD1. Crowding-Out Effects

Panel A. Excluding whole OECs with homogeneous real-world question				
	(1)	(2)	(3)	(4)
Treated=1	-0.130***		-0.130***	
	(-10.471)		(-10.503)	
Text=1		-0.061***		-0.062***
		(-3.756)		(-3.772)
Text # Image=1		-0.168***		-0.167***
		(-5.818)		(-5.817)
		10,292 observations		
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Batch FE	Y	Y	Y	Y
Platform FE	Y	Y	Y	Y
Question FE			Y	Y
Panel B. Excluding firm-questions asked later than arrival of homogeneous real-world question				
	(1)	(2)	(3)	(4)
Treated=1	-0.130***		-0.130***	
	(-10.479)		(-10.510)	
Text=1		-0.061***		-0.062***
		(-3.759)		(-3.774)
Text # Image=1		-0.168***		-0.167***
		(-5.822)		(-5.821)
		10,294 observations		
Controls	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y
Batch FE	Y	Y	Y	Y
Platform FE	Y	Y	Y	Y
Question FE			Y	Y

In this table, we replicate our baseline results using two variant samples. All specifications remain the same as in Panel A of Table 6, except that we exclude firm-questions from OECs with questions that are the same as ours from other OEC participants in Panel A, and we exclude firm-questions we asked later than the arrival of such homogeneous question on such OECs in Panel B.

Table OD2. Responsiveness to AI-Styled Questions of Bustling Managers

	(1)	(1)
Indicator:	Bustling Manager	Longer Duration
Indicator # Treated=1	-0.046	-0.054**
	(-1.439)	(-2.036)
Indicator=1	0.017	0.098***
	(0.759)	(4.973)
Treated=1	-0.121***	-0.111***
	(-8.632)	(-7.218)
Observations	10344	10348
Controls	Y	Y
Industry FE	Y	Y
Batch FE	Y	Y
Platform FE	Y	Y
Question FE	Y	Y

In this table, we run Probit regressions of Response on the interaction of Treated with an indicator. The interacted indicator is Bustling Manager and Longer Duration in columns (1) and (2), respectively. Treated equals 1 for treated firm-questions, and 0 otherwise. Bustling Manager equals 1 when # Q&As is among top 25%, and 0 otherwise. Longer Duration equals 1 when duration of the OEC is above sample median, and 0 otherwise. All other specifications remain the same as in Table 7. Table Variable Definitions provides detailed variable definitions.