

Bot-induced social media manipulation and stock market distortion

Daniel Bradley
University of South Florida
Email: danbradley@usf.edu

Jan Hanousek, Jr.
University of Memphis
Email: jan.hanousek@memphis.edu

Jan Hanousek
Mendel University in Brno, and CEPR London
Email: jan.hanousek@mendelu.cz

Dominik Svoboda
University of Memphis
Email: dsvoboda@memphis.edu

December 2025

Abstract

Using an exogenous shock to bot activity, we show that bot presence and their impact rise with the ease of infiltrating social platforms. Bots significantly manipulate social media attention and are most likely to target firms that are more volatile with recent momentum and around earnings releases. When bot penetration is high, stock price synchronicity increases and post-earnings announcement drift is stronger, indicating that bots lead to less efficient prices. Bot accounts promoting stocks generate sharp positive returns, followed by partial reversals and retail order imbalances, likely exacerbating trading frenzies. Overall, bots create information frictions that distort market prices. Our evidence reinforces regulators' concerns about AI-driven stock manipulation.

Keywords: Information frictions, investors' attention and sentiment, post-earnings announcement drift, social media bots, stock market distortion, stock price synchronicity, trading frenzies.

JEL Codes: G11, G14, G30, D22, C26.

Bot-induced social media manipulation and stock market distortion

“Stocktwits is a bot factory, and overrun with paid shills who farm engagement to sway social sentiment.”¹

1. Introduction

The intersection of social media and stock market research has grown exponentially over the past several years. Numerous papers analyze platforms such as *StockTwits* (Cookson et al., 2024), *Twitter* (Bartov et al., 2018), *Seeking Alpha* (Farrel et al., 2022), and *Reddit* (Bradley et al., 2024). This research has frequently been at odds with early work, which overwhelmingly implied retail investors were uninformed noise traders. Analysis of investor opinions across these retail-dominated social media platforms suggests that, in many instances, retail trades are indeed informative. In fact, according to a survey by Bloomberg Intelligence, nearly all hedge funds now actively monitor retail activity, which was not the case less than a decade ago.² Researchers have focused on two main measures on social media platforms to gauge interest in a particular stock and the general aggregate mood – attention and sentiment. An important study by Cookson et al. (2024) provides direct comparisons of sentiment and attention across platforms, noting that while attention is generally correlated, sentiment varies widely.

The presumption across this social media literature is that these measures capture human communication and interaction. However, bots, or software robots designed to execute tasks autonomously, have become pervasive on social media platforms. For instance, during the purchase of Twitter (now X), Elon Musk commissioned a study (Duffy and Fung, 2022) that estimated that roughly 11% of accounts on the platform were bots. This number is likely higher. Hanousek et al. (2025) find that bots account for 17% of engagement activity on Twitter. The fear over bots on social media platforms arose on Twitter due to suspected bot involvement during the 2016

¹ Reddit poster:

https://www.reddit.com/r/Superstonk/comments/ut7ujm/stocktwits_is_a_bot_factory_and_overrun_with_paid/

² See: <https://tokenist.com/85-of-hedge-funds-now-use-jpmorgans-tool-to-monitor-retail-on-social-media/>

presidential election (e.g., Bessi and Ferrara, 2016). Regulators’ concerns have broadened from bots’ influence on election interference and other political dealings to the potential harm they pose to children.³ Ultimately, the fear is that bots can significantly manipulate people’s attention and sentiment, leading to adverse decision-making.

If bot activity across social media platforms is pervasive but not identified, it would likely lead to biased measures of attention and sentiment. This is also true in the case of stocks. For instance, take the common measures of firm-level attention and sentiment as in Cookson et al. (2024). Those measures could include bot-written messages, which would introduce bias into these common measures used by researchers and market participants, such as hedge funds. In particular, if one actor is employing bots to manipulate their position, measures of sentiment and attention would not truly reflect the wisdom of the crowds (Chen et al., 2014). If multiple bot networks coordinated their efforts, this would exacerbate the bias.

In our study, we examine the impact of bots on social media platforms and their effect on price efficiency. We take two approaches to do this. We first rely on plausibly exogenous shocks to bot activity. As described in more detail in Section 2.2, we exploit technological innovations in efforts aimed at restricting bot activity. These innovations are unrelated to economic conditions, traffic on social media platforms, and impact all social media platforms simultaneously. These attractive features allow us to make causal inference. Specifically, we exploit periods when bot access is easy versus more difficult. We find that classically defined measures of attention are significantly affected by the presence of bots. When bot access is easier, attention significantly increases. Further, social media appears to function differently during periods of easier bot access. Messages tend to be more urgent, and attention is far more persistent than during periods of harder bot access.

³ In a September 2025 hearing, parents allege that bots caused their children’s suicide:
<https://www.techpolicy.press/transcript-us-senate-hearing-on-examining-the-harm-of-ai-chatbots/>

After demonstrating that bots can impact attention, we turn our attention to their potential impact on capital markets. Past research (i.e., Bradley et al., 2024) shows the crucial role of social media in sharing and obtaining information. If a large number of messages are bot-induced, this can have two effects. First, it can overexaggerate attention on stocks that are targeted by bots. Second, given that investors have limited attention, excessive attention to the set of targeted stocks will drive attention away from other stocks that otherwise would have received it. Both of these effects can increase noise and drown out new information. As discussed, when bots artificially exaggerate attention on some stocks, this may cause investors to overlook non-targeted bot stocks. In such cases, prices may not reflect available idiosyncratic attention and have greater co-movement with the market and the industry (Chen et al., 2023; Goldstein, 2023). To test this, we examine price synchronicity, the degree to which an individual stock's price moves with the overall market. When price synchronicity is high, it does not reflect idiosyncratic information. This is precisely what we find—price synchronicity is higher on days with easier access to social media by bots. This suggests that bots are negatively impacting the information environment, causing less informative prices.

Next, we examine the impact of bot activity on post-earnings announcement drift (PEAD)—the market anomaly where firms' stock prices slowly react to the surprise in firms' earnings announcements. DellaVigna and Pollet (2009) find PEAD is stronger on earnings announced on Fridays. They argue investors are distracted on Friday's as they look forward to the weekend. Similarly, Hirshleifer, Lim and Teoh (2009) find that when more firms issue earnings on the same day, the initial market reaction is weaker and PEAD is stronger. Investors have limited attention and other firms' earnings distract their attention. In our setting, we predict that bots have the same ability to distract attention. We find that when bot access to social media platforms is difficult, investors are subject to less distraction and firms' unexpected earnings are immediately incorporated into stock prices with no drift. However, when bots have an easy time infiltrating social media platforms, we

find the initial reaction is incomplete and significant drift that persists for 40 trading days post-earnings announcement. Taken together, these findings support the view that bots induce information frictions.

In the second part of the paper, we take a different approach to examining the impact of bot activity on capital markets. Here, we explore the content and behavior of accounts identified to be bots on *Stocktwits*. We identify a total of 756 accounts, which is trivial in terms of the number of accounts on the platform. However, these accounts generate 0.6% of *all* the 800 million messages on the *Stocktwits* platform. Thus, we are confident that accounts identified as bots are indeed bots. We first examine the characteristics of firms that bots are likely to target. We find they are larger, more volatile, have more momentum and are more likely to target stocks around quarterly earnings announcements.

Before we tackle the impact of actual bot activity on stock prices, we first validate our empirical strategy using vulnerabilities to bot restriction (i.e., bugs). In periods where low bug activity is reported, this would coincide with low bot penetration. Conversely, when high bug activity is reported, bot activity should be elevated. Using actual bot accounts described above, this is precisely what we find. Further, this test should also reinforce the view that the accounts identified are bots, indeed are bots.

When bots target a specific stock, what does this do to stock prices? When a bot promotes a stock, we find CARs are 1.16% the day it is targeted. Over the next 20 trading days, we find a partial reversal of approximately one-third of this return. This finding suggests that bots can distort market prices.

Our last tests focus on retail order imbalances and retail trading frenzies. We find that when a stock is promoted by a bot, it is positively associated with significant retail order imbalances. For instance, retail order imbalances are 8.3% higher on the day a bot promotes the stock. This imbalance

persists beyond the target date for at least 5 days. Given that retail order imbalances are related to retail trading frenzies, our results suggest that bots have the potential to amplify trading frenzies. We empirically test this conjecture using the definition of trading frenzies in Green and Jame (2025). We find that bot promotion is elevated, and likely exacerbates, retail trading frenzies.

Our paper contributes to at least three broad areas of research. First, we contribute to the nascent literature on social media and stock markets. Studies have analyzed various social media platforms independently such as *Twitter* (Blankenspoor et al., 2014; Bartov et al., 2018; Gu and Kurov, 2020), *Seeking Alpha* (Chen and Hwang, 2022; Farrel et al., 2022; Kogan et al., 2023), and *Reddit* (Bradley et al., 2024). Most of these studies focus on capturing measures of investor attention and sentiment. Cookson et al. (2024) considers multiple social media platforms and evaluates them simultaneously to get an aggregate ‘social signal.’ While our paper mostly focuses on *Stocktwits*, we also consider multiple social media platforms. We diverge from the existing literature by considering the impact of bots on social media signals. Our results suggest they can have profound effects on social media measures employed in finance research.

We further contribute to the literature on retail investors. Retail investors rely on social media for information and investment advice (Bradley et al., 2024; Cookson et al., 2024). Recent evidence suggests that retail trades may be informed (Kelly and Tetlock, 2017; Boehmer et al., 2021; Farrell et al., 2021). Given that close to half of retail investors use social media to act on financial advice, our results suggest that these investors may unknowingly be manipulated into making decisions they otherwise would not have.⁴ Our paper also contributes to work on retail trading frenzies (Goldstein et al., 2013; Green and Jame, 2024). We find bots that promote stocks are associated with retail order imbalances, which have the potential to exacerbate retail trading frenzies.

⁴ https://www.betterment.com/hubfs/PDFs/b4c/2024%20Retail%20Survey_Betterment.pdf

Finally, our paper contributes on the evolving literature on the effects of AI on the financial markets and investor behavior. Prior to AI technology proliferation, bots were limited in their functionality. Now, AI-powered bots and bot networks (Yang et al., 2019) are far more advanced, and capable of much more complicated tasks appearing more human-like. Dou et al. (2025) develop a theoretical model where they outline how AI-powered trading agents can be capable of coordinating, without prior communication and without detection from other trades. Their models suggest how AI can be misused for market manipulation. Our results build on their model to outline how AI can coordinate on social media. In our setting, a single AI agent can be put in control of a large number of accounts on any social media platform and use them for coordinated manipulation.⁵

Our paper has important policy implications. Our results indicate that bots and AI-powered bot networks constitute a serious threat to the integrity of social media and their roles, as they distort information and create information frictions. Further, our paper shows that this manipulation distorts stock market prices and leads to price inefficiency. Our evidence supports regulators' concerns regarding the potential harm AI may cause to financial market integrity.⁶

This paper is organized as follows. We provide a history of bot accounts and bot detection techniques in Section 2. We describe the data in Section 3. We provide an overview of theoretical models that can be used to understand how bots can impact information distortions and frictions in Section 4. Section 5 reports results on the impact of bot activity and price synchronicity and post-earnings announcement drift. Section 6 identifies bot accounts and provides results on how bots impact individual firms targeted by them. Section 7 concludes.

2. Bot detection

⁵ Unlike the potential collusion outlined in Dou et al. (2025), manipulation of social media does not require large capital outlays. Hanousek et al. (2025) document that buying engagement on Twitter can cost less than five dollars per 100 interactions.

⁶ <https://www.axios.com/2023/08/12/artificial-intelligent-stock-market-algorithms>

2.1 Social media bots

Social media use has grown significantly over the past decade, with more than half of the world's population actively using at least one platform.⁷ Moreover, at least 30% of Americans turn to social media for financial advice.⁸ In theory, social media can serve as a public forum, where users can share and receive information with their peers, potentially leading to better financial decision-making (e.g., Ouimet and Tate, 2020; Goldstein et al., 2025). However, social media bots—accounts that are either partially or fully controlled by software designed to complete specific tasks—are pervasive on social media platforms (Hanousek et al., 2025). They solely exist to shape public perception or push a certain narrative (Yang et al., 2019; Keller et al., 2020). Bots can not only distort the dialogue, but they also introduce noise and information frictions.

While bots have existed on social media platforms since their launch, perhaps the first indication of bot prevalence on Twitter was released in 2009, which found that 24% of accounts are bots.⁹ One of the first high-profile bot accounts on Twitter that attracted the media's attention was @Horse_ebooks. The publication *Wired* wrote, 'One of the funniest people on Twitter isn't a person at all. It's a bot called @Horse_ebooks.' However, it wasn't until the 2016 election of Donald Trump that political disinformation from bot activity became headline news, subsequently drawing concern among regulators (e.g., Bessi and Ferrara, 2016; Ferrara, 2017).¹⁰

2.2 Bot detection through Captcha

The primary tool for combating bot presence on any online platform is the Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA). Providers, such as Cloudflare, offer services that monitor internet traffic for websites, issuing a CAPTCHA challenge

⁷ See <https://www.demandsage.com/social-media-users/>.

⁸ <https://www.bankrate.com/investing/financial-advisors/americans-financial-advice-top-place/>

⁹ <https://cacm.acm.org/news/researchers-se-social-bots-to-expand-twitter-networks-and-influence-people/>

¹⁰ There were multiple Senate Select Committee on Intelligence meetings in 2017 and 2018 regarding political interference of bots.

upon detecting suspected bot activity.¹¹ For instance, a user may be given a variety of pictures and asked to click on all the ones that feature a motorcycle to prove they are human. Bot operators work to subvert these new techniques and develop proprietary methods to solve CAPTCHA algorithmically. This creates a continuous technological arms race, a game of cat and mouse, analogous to the battle between antivirus software and computer viruses.

The application of CAPTCHA technology is very narrow and does not disrupt human activity on social media. Importantly, technological shocks to CAPTCHAs occur randomly and are exogenous to investor discussions. This allows us to utilize them to study the impacts of bot activity and presence on social media. The challenge in identifying technological shocks to CAPTCHA is that the CAPTCHA providers do not provide an exhaustive list of all improvements made. As a result, we adopt the approach described below.

We rely on publicly available information on GitHub (www.github.com), which is the largest developer platform for storing open-source code and programs. The platform serves as a repository for developers to store code and programs for others to use. While the majority of codes and programs on GitHub are open source (i.e., free to use), many providers upload parts of their code necessary to employ their programs by paying customers, and use the platform to respond to questions and reports. Proprietary software developers often use and rely on code from GitHub and other open-source repositories to cut down on cost (Hoffmann et al., 2024) and is frequently used by developers to showcase their work (Wright et al., 2023).

Each application on GitHub is contained within its own repository, which includes full documentation, a link to each code used by the application, as well as a list of updates and bugs reported by the application's owner and users. We collect data on all repositories on GitHub that are focused on CAPTCHA—implementing, improving, or providing an alternative to it. There are 1,595

¹¹ See <https://www.cloudflare.com/application-services/products/turnstile/>.

repositories fitting this description. Next, we parse these GitHub repositories to identify issues (i.e., bugs) reported and resolved in each repository. We focus on bugs due to their correspondence with technological evolution and how they can accurately describe current conditions. For instance, in a period when many individuals are reporting bugs to these repositories, it is because CAPTCHA services are either not working or not functioning properly. During these periods, bots should have easier access to the platforms, as CAPTCHA will not be operating effectively and thus will be less able to block their activity. Conversely, in periods with very few bugs reported, CAPTCHA should be working as intended, and therefore, bots have a harder time accessing the platforms.¹²

To capture the potential long-term consequences of such bugs, we define our primary variable of interest, *CAPTCHA bugs*, as the natural logarithm of one plus the number of unsolved reported issues (bugs). Unsolved bugs is a proxy for the potential long-term vulnerabilities that bots can exploit. When a developer resolves the problem, the potential for bots to exploit any platform decreases and thus consequently does our measure.¹³

3. Data

Our series for identifying bot activity is from 2016 to 2023. The beginning period is when social media bots received heightened attention in the popular press and drew regulators' concerns following the 2016 Presidential election. We stop in 2023 because the traditional CAPTCHA has been rendered largely ineffective by AI and replaced.¹⁴ Appendix A provides a time series plot of

¹² Consider the case of a new computer virus. When launched, the risk to software users is the highest. The lag between its launch date and when it is identified, the time it takes for developers to create a solution to prevent it from infiltrating devices, and how many users and how long it takes them to update their virus protection software will determine the effectiveness of the virus. If we could identify when a solution was found and we could further determine the number of users who have updated their antivirus software, we could create a similar measure for high versus low virus vulnerability that would be analogous to our bot detection method.

¹³ Hanousek et al. (2025) demonstrate that CAPTCHA bugs significantly predict bot activity on Twitter, providing validation for using this measure as a proxy for bot activity.

¹⁴ See, e.g., <https://blog.cloudflare.com/end-cloudflare-captcha/>

CAPTCHA bugs. The series is quite volatile, suggesting that developers and bot operators are continually evolving.

Our primary social media data focuses on *StockTwits*. We collect message-level data and use VADER (Hutto and Gilbert, 2014) to obtain message-level sentiment.¹⁵ The benefit of *StockTwits* in our setting is that it is possible to collect *all* the messages on the platform. However, we also employ *Twitter* and *Seeking Alpha* to account for their sentiment and attention, as outlined by Cookson et al. (2024). For *Twitter*, we obtain sentiment and attention metrics from Bloomberg (Bartov et al., 2018), and we collect article-level sentiment (specifically ESS) for *Seeking Alpha* from RavenPack.

Unlike Cookson et al. (2024), who focus on cross-sectional analysis, our analysis centers on the time-series dimension. As such, instead of focusing on the top 1,500 firms by the number of *StockTwits* messages posted over the entire sample, we instead use the top 1,500 firms each day to avoid look-ahead bias. Further, as described below, we adapt their measure of attention.¹⁶

Cookson et al. (2024) calculate attention as:

$$Attention_{i,t} = \frac{Messages_{i,t}}{\sum_i Messages_{i,t}} \quad (1)$$

where $Messages_{i,t}$ is the number of messages of stock i over time t . Note that bots can influence attention in two opposing ways. They can actively increase attention for targeted stocks in the numerator, which can lead to information distortion (e.g., Kogan et al., 2023), while simultaneously causing a relative decrease in attention for non-targeted stocks by inflating the overall message count in the denominator, thereby inducing information frictions (e.g., Hirshleifer et al., 2009). Given that bots can both amplify and suppress the traditional attention metric, we adopt a refined definition to better isolate their effect. Specifically, we define attention as $\log(1 + Messages_{i,t})$. This method ensures that bot activity can only influence attention in a single direction.

¹⁵ The data can be found here <https://github.com/Jaxingjili/StockTwits-Data-Processing-and-Analysis>.

¹⁶ As described later, we also report results using equation 1 and strictly follow Cookson et al. (2024).

We define control variables following Cookson et al. (2024). First, we collect information on attention and sentiment from the *Dow Jones Newswire* from Ravenpack. We keep all articles with a relevance score above 75 and utilize the Ravenpack Event Sentiment Score (ESS), which aggregates the article-level sentiment to the firm-day level by averaging firm-specific sentiment across articles within a given day. We define daily abnormal returns as the difference between the firm's daily return and the value-weighted market return, using CRSP data. Firm-level controls include size, defined as the natural logarithm of market value of equity; volatility, defined as volatility in returns over the past five days; lagged market abnormal returns, defined as daily cumulative abnormal returns relative to the value-weighted index over periods (t-1) to (t-5) and (t-6) to (t-30); and indicators for 8-K filings and earnings announcement days.

4. Social media platform manipulation: Theoretical considerations and predictions

4.1 Information distortion

How do bots impact dialogue between investors, what information they share, and how they process that information? Mechanically, bots can influence investor perception primarily by artificially increasing the frequency of stock mentions, thereby creating a potentially misleading signal of popularity regarding both attention and collective sentiment (Hanousek et al., 2025). Beyond simple inflation, bots can also directly worsen information quality by introducing incorrect or fabricated messages (Kogan et al., 2023). This dual role of bots, inflating volume and potentially corrupting content, increases noise and could lead to significant information frictions.

Panel A of Table 1 presents the analysis of the relationship between *CAPTCHA bug* severity, a proxy for the ease of bot entry, and social media attention. In this analysis, we use the number of messages on both *StockTwits* and *Twitter* platforms, respectively. We observe that exogenous shocks to CAPTCHA technology significantly impact attention levels. For example, a 1% increase in reported CAPTCHA bugs is associated with a 0.32% increase in messages on *StockTwits*. Likewise, a 1%

increase in reported CAPTCHA bugs is associated with a 0.27% increase on *Twitter*.¹⁷ This result suggests that bot activity on social media platforms is likely substantial and has the ability to systematically influence public discussion.

Insert Table 1

In Panel B of Table 1, to show robustness, we also use the attention metric defined by Cookson et al. (2024) discussed earlier. We employ their publicly available first PCA component of attention and replicate their data for attention across all three social media platforms (*Stocktwits*, *Twitter*, and *Seeking Alpha*). We employ the raw and standardized versions of their measure. While their measure is constructed from 2012 to 2021, we further restrict the sample to begin in 2016, when concerns about bot manipulation surrounding the 2016 Presidential election received significant media and regulators' attention.¹⁸ To facilitate direct comparison, we focus on the same sample of firms, namely the top 1,500 most-discussed firms over the period 2012-2021. The first model presents the first PCA attention factor created by Cookson et al. (2024), which covers all three platforms. We again see that *CAPTCHA bugs* is positive and significant. When examining *Stocktwits* and *Twitter* separately, we find similar results. These results further bolster the claim that bot activity is a powerful predictor of attention.

While Table 1 focuses exclusively on the impact of bots on social media attention via increased message volume, bot activity also fundamentally affects market sentiment through the content of its messages. Crucially, this information distortion is not limited to retail investors. The model developed by Goldstein et al. (2025) demonstrates an equilibrium where sophisticated, well-informed individuals receive and trade on information derived from coarsely informed investors, even if it leads to

¹⁷ Note that while Cookson et al. (2024) also examine *Seeking Alpha*, we do not here because this platform is characterized by highly selective contributors, low message frequency, and a traditional media structure. This platform is less susceptible to potential bot-driven manipulation than *StockTwits* and *Twitter*. However, we do examine *Seeking Alpha* in some subsequent tests.

¹⁸ The results are robust if our sample completely mirrors their 2012-2021 sample.

suboptimal outcomes. This framework explains why hedge funds collect and utilize social media data in their models. However, bot detection is notoriously challenging (e.g., Broniatowski et al., 2018), so the information distortion potentially caused by bots extends even to large institutional investors.

Consequently, bots can serve as a potent tool for information manipulation, analogous to fake news analyzed by Clarke et al. (2020) and Kogan et al. (2023). This risk is compounded by advancements in AI, as highlighted by Goldstein et al. (2025), who illustrates how AI-powered trading algorithms can coordinate without prior communication or detection to manipulate prices. A similar framework could be implemented on social platforms, with a single AI agent managing an entire bot network.¹⁹

Insert Figure 1

In Figure 1, we present the word cloud of the most popular words associated with the periods when bots have easier access to the platform (i.e., when CAPTCHA bugs are in the upper quartile), compared to days with harder bot access on the platform. The words indicate that the messages are bullish, urgent, and emotional. This is precisely what can be used to distort information and impact on how investors perceive the company.

4.2 Distractions and information friction

The previous section documented the capacity of bots to distort information regarding specific targeted stocks. However, their very presence introduces broader information frictions by altering the equilibrium distribution of attention across the platform. The systematic promotion of certain stocks, or even the generation of increased noise by bot networks, shifts investor focus. The algorithmic

¹⁹ While not the focus of our study, AI-powered reinforcement learning would be ideal for this purpose. An AI agent could be responsible for hundreds or even thousands of bot accounts and instructed to focus discussions on specific stocks at certain times. In the reinforced learning setup, it would be rewarded for every positive interaction with a user outside its network and punished for being detected by the platform or accused/suspected by other users.

nature of social media platforms, often using reverse-engineerable methods (Chakrabarti et al., 2023), could further intensify this dynamic and redirect investors' attention.

Insert Table2

Table 2 provides a comparative analysis of message characteristics. Specifically, we compare message volume, length, and active user count across platforms under varying conditions of bot entry accessibility. In particular, we define easy (hard) bot access as periods when CAPTCHA bugs are in the top (bottom) 25th percentile across our sample period. Consistent with the predictions of information friction theory, we observe that periods facilitated by easier bot access are associated with over a 60% increase in active users (23.4k vs. 15.7k for easy versus hard bot access, respectively). Likewise, we observe an approximately 50% increase in the number of messages (112.2k vs. 73.9k for easy versus hard bot access, respectively), as well as a modest increase in message length. This systemic alteration of the communication environment associated with easier bot entry is likely to increase information frictions. These information frictions may result in systematic overreaction of market participants' assessment due to bot-driven promotion and underreaction to firms not receiving this attention as the artificial attention distracts investors away from non-targeted stocks.

To understand the systemic impact of targeted bot campaigns, we analyze how the presence of bots distorts the distribution of social media attention. Figure 2 illustrates this result and documents the persistence of highly popular stocks across days characterized by differential bot accessibility.

Insert Figure 2

Even when we standardize message volume to control for platform activity, attention exhibits significantly greater persistence during periods of low bot entry barriers. This tendency for bots to systematically promote stocks for an extended duration generates increased platform noise, which is expected to intensify information frictions. We provide a more comprehensive analysis of attention distribution.

Insert Figure 3

In Figure 3, we use the definition of attention provided by Cookson et al. (2024), as described in Section 3. This approach intuitively takes into account the potentially limited attention of investors (i.e., DellaVigna and Pollet, 2009) and thus makes it ideal to observe how limited attention of investors is distributed when there are more or fewer barriers to entry for bots. The observed patterns suggest that the distribution is bimodal. The distribution of attention during days with difficult bot access indicates that more stocks achieve moderate attention. Conversely, during periods of easier bot entry, more extreme observations are concentrated around zero or shifted to the right. Thus, when bots are active, a small number of stocks receive the majority of attention.

Overall, these stylized facts suggest that bots can exert a powerful influence on social media, leading to a distortion in attention allocation that precipitates increased information frictions. This mechanism, in turn, is expected to result in potential market underreactions to genuine news that is released when bot activity is elevated. This conceptual relationship is highly analogous to the investor inattention model proposed by DellaVigna and Pollet (2009), outlining how investor inattention due to distractions may lead to underreaction to news. To empirically validate this hypothesis, we examine the effect of bot activity on post-earnings announcement drift (PEAD) (Ball and Brown, 1968) and price synchronicity (Chen et al., 2007).

5. Bots and information frictions and distortions

5.1 Bots as drivers of information frictions

In this section we examine how bots may potentially induce information frictions and stock market distortions. Section 5.1.1 focuses on bot activity and price synchronization and section 5.1.2 examines bot presence and post-earnings announcement drift (PEAD).

5.1.1 Price synchronicity

As discussed in Section 4.2, bot activity can lead to substantial information frictions, which may negatively impact the market's informational efficiency. When bots cause investors to overlook stocks not promoted by bots, it may lead to underreaction, and the price may not reflect available idiosyncratic attention. In such cases, the stock price should have greater co-movement with the market and the industry (Chen et al., 2023; Goldstein, 2023). To test this hypothesis, we analyze price non-synchronicity during periods of high and low bot activity on social media. When price non-synchronicity is high, markets face information frictions. In our setting, we conjecture that high bot activity is associated with lower levels of price non-synchronicity. To test, we run the following regression model:

$$Ret_{i,t} = \alpha_i + \beta_{1,i}Market\ return_t + \beta_{2,i}Industry\ return_t \quad (2)$$

where *Market return_t* is defined as the value-weighted return from CRSP and *Industry return_t* is the return of the corresponding Fama-French 48 industry portfolio. Price non-synchronicity is $1 - R^2$ of the resulting regression. We run this regression separately for each firm and separately for periods with low and high bot activity, defined as the top and bottom 25% of *CAPTCHA bugs*.

Table 3

Table 3 reports the results. When bot access is more difficult, price non-synchronicity is 0.73 compared to 0.66 when bot access is easier. This difference is highly statistically significant.²⁰ This implies that when bots have easier access to the platform, the market and industry index explains more variation in firms' stock returns. However, when bots have more difficulty accessing platforms, stock returns reflect more idiosyncratic information (Chen et al., 2007). These empirical patterns are consistent with our hypothesis that bots decrease informational efficiency by introducing information frictions. We next examine the impact of bots on post-earnings announcement drift (PEAD).

²⁰ Testing the difference between $1 - R^2$ in two independent samples is equivalent to testing the equality of residual variances. This leads to a standard F-test, with the underlying F-statistic being $0.73/0.66 = 1.106$ and the p-value is of order $2e-6$.

5.1.2 Post-earnings announcement drift (PEAD)

Post-earnings announcement drift is a well-documented market anomaly, first identified by Ball and Brown (1968), which remains an elusive puzzle that continues to interest researchers. It is the tendency of a stock’s price to continue moving in the same direction as an earnings surprise for a period of time after the company’s earnings announcement. DellaVigna and Pollet (2009) specify a theoretical model and empirically demonstrate that there is significantly stronger post-earnings announcement drift (PEAD) on earnings announced on Fridays, as investors are distracted by weekend plans and activities. In our context, we conjecture that high bot activity will introduce frictions to the information environment and be associated with higher levels of PEAD. In particular, when bots concentrate their efforts on a select number of stocks, thereby amplifying attention on these stocks, market participants place less attention on non-targeted stocks than they would otherwise. These market frictions cause inefficiencies and information to slowly diffuse into market prices.

Following the extant literature, we compute abnormal returns following quarterly earnings announcements and how they are impacted by standardized unexpected earnings (SUE).²¹ We use the following regression specification:

$$DGTW\ returns_{i,t} = SUE_{i,t} + Attention_{i,t} + Sentiment_{i,t} + X_{i,t} + Day_t + Industry_i, \quad (3)$$

where $DGTW\ returns_{i,t}$ corresponds to the DGTW-adjusted return (Daniel et al., 1997), consistent with prior literature on PEAD (e.g., Ben-Rephael et al., 2017), $SUE_{i,t}$ is standardized unexpected earnings, and $Attention$ and $Sentiment$ refer to the attention and sentiment measures of stock i at time t on each of the social media platforms: *StockTwits*, *SeekingAlpha*, and *Twitter*. Following Cookson et al.

²¹ Consistent with the literature (e.g., Ben-Rephael et al., 2017; Liu et al., 2023), we define SUE as the actual EPS reported minus the median EPS forecast from the analyst, scaled by the share price at the time of the announcement.

(2024), the vector of control variables $X_{i,t}$ include volatility, past returns, and indicator variables for firms' 8-K filing and earnings release dates. All our regressions include day fixed effects to absorb other events occurring on days other than bot activity days. We also include the Fama-French 48 industry fixed effects in this specification.

Insert Table 4

Table 4 presents the results. To determine the role of bot activity on PEAD, we analyze the effect of SUE on cumulative returns separately for periods with easier and harder bot access. During periods of harder bot access, on the day of earnings announcements ($DGTW\ return\ (t)$), the coefficient on SUE is positive and statistically significant at 0.069 compared to an insignificant 0.019 when bot access is easier. This difference, and primarily the difference in statistical significance, suggests that when bot activity is restricted, earnings surprises are immediately incorporated into the stock price, whereas when bots have easier access, we observe persistent and long-lasting effects of SUE. Looking beyond $DGTW\ return\ (t)$, we observe the opposite pattern. During hard bot access, the drift does not last past day +1. However, when bots have easier access to the platforms, the drift remains significant, up to 40 trading days past the earnings release. Recall that the specification includes a day fixed effect, which controls for other unobservable effects and events occurring on days with higher or lower bot activity.

To further visualize the effects of bots on PEAD, we follow Ben-Rephael et al. (2017) and plot bot activity for positive and negative SUE. For each group, we calculate the group's conditional return by multiplying the SUE average (conditional mean) by the group's relevant SUE regression coefficient as estimated in Table 4. As shown in Figure 4, the most apparent visual is the negative SUE high bot activity group. One plausible explanation for this result is that bots tend to be promoters and thus are overly optimistic of the firms they target. Thus, companies with negative SUE are less likely to be targeted by bots and attention is shifted elsewhere. Nonetheless, we find some evidence of

PEAD for positive SUE, but economically the effect is much smaller and only statistically significant over longer time horizons. In either case, what is clear is that when bot activity is restricted, PEAD is as well.

Insert Figure 4

Collectively, this section shows that shifting investor attention to bot-targeted stocks causes inattention in non-targeted stocks. As a result, our empirical evidence suggests that price synchronicity and PEAD are more likely to exist when bot access is easier and distracts attention away from stocks that otherwise would have received it.

6. Identification of bots and information distortion

In the previous section, our identification relied on the ability of bots to infiltrate social media platforms. The appealing feature of this method is that it is presumably exogenous to stock markets and investment behavior and simultaneously impacts every social media platform. However, it is unable to isolate *actual* bot activity on individual stocks. In this section, we take a different empirical approach to examine the impact of bots. Specifically, we attempt to identify bot accounts and examine their behavior.

Bot accounts are notoriously difficult to detect. However, users of *StockTwits* appear to be aware of the bot problem, referring to bot accounts as “pumpers” because they primarily promote bullish sentiment.²² While bot algorithms exist, many argue that they are unreliable as bot accounts have become more sophisticated.²³ Generally, only more experienced users and those with better

²² For instance, consider this Reddit post: ‘stocktwits is full of pumpers and unsophisticated investors that just listen to someone with more followers without doing their own diligence. their king seems to be some guy named turtle. whatever he invest in, his followers pump and he dumps. its quite sad. be careful’;
https://www.reddit.com/r/pennystocks/comments/lfpecy/watch_for_pumps/

²³ Twitter has directly warned against the use of publicly available bot detection techniques due to their unreliability and reliance on only publicly available data. The detailed discussion is available at
https://blog.twitter.com/en_us/topics/company/2020/bot-or-not.

reasoning abilities can identify them (Kenny et al., 2024). Thus, rather than using an algorithm, we instead rely on bot accusation of other users to detect bot activity.

Using our data collected on *StockTwits*, which spans over 800 million observations, we analyze the text to identify bot accusations and manually verify each accusation. We flag all messages mentioning the word 'bot' or 'bots' and manually evaluate whether the message is, in fact, an allegation of being a bot made by another user in a reply. We then require that the accused account not respond to the accusation and remain active on the platform. Our final sample contains 837 valid bot accusations, out of which 756 accounts remained active following the accusation. These accounts generated over 4.6 million messages, corresponding to an average of more than 1,600 messages per day. Out of these messages, over 2.6 million mentioned a single stock, which serves as a filter for analyzing sentiment and attention, following Cookson et al. (2024). Thus, while this small number of sample accounts represents a lower bound for bot activity on the platform, they represent an astonishing 0.6% of *all* activity on the platform. Their activities also cover over 6% of stock-days in our sample, corresponding to more than 168,000 observations, which occurred following a day when the stock was positively mentioned by a bot. Thus, we are fairly confident that these accounts are correctly identified as bot accounts.

6.1 Validation of bot accounts and CAPTCHA as identification

Our empirical strategy in section 5 exploited exogenous vulnerabilities in bot penetration on social media platforms. The identifying assumption is that these vulnerabilities restrict bot access. We validate this assumption in this section by directly estimating bot activity relative to the difficulty for bots to access social media platforms. Recall our proxy for this difficulty is *CAPTCHA bugs* where high values of bugs indicate periods where the technology was not functioning properly, leading to easier bot access. Using accounts identified as bots above, we test if these accounts are more active

when bugs are more frequent. Not only does this test help validate our identifying assumption, but it should further enforce the view that these accounts are indeed bots.

Insert Table 5

In Table 5, we regress the number of messages (column 1) and the natural log number of messages on *Stocktwits* on *CAPTCHA bugs*. Additionally, we include an account fixed effect, which absorbs unobservable time invariant variables. Standard errors are clustered at the account and day level. Both the number of messages and the natural logarithm of number of messages significantly increase with *CAPTCHA bugs*. Economically, a 1% increase in *CAPTCHA bugs* is associated with 10 more messages per bot account, which corresponds to an increase of 0.7% more messages per account. These results help validate our identification strategy in section 5 and should alleviate concerns about our identification of bot accounts. We now turn our attention to the impact of bot accounts on stock market outcomes.

6.2 Do bots target firms with certain characteristics?

We first examine if bots target firms with characteristics that differ from the average firm. We focus on the usual suspects: size, volatility, and momentum. We conjecture that bot operators are more likely to target firms they can more easily manipulate and exploit. This would imply smaller, more volatile firms with more momentum. Likewise, we predict bots target firms around information events such as quarterly earnings announcements. Table 6 presents these results.

Insert Table 6

Interestingly, firms targeted by bots are larger than those not targeted by bots. This is not what we expected, but perhaps in larger firms that have more messages posted, bots are more easily disguised. However, we indeed find that firms that are more volatile and have more price momentum are more likely to be targeted by bots. For instance, bot-promoted stocks have almost double the volatility (5.3% compared to 3.6%) and a factor of 10x momentum. CARs on days $(t-30, t-6)$ are

3.35% for bot-targeted stocks relative to 0.30% for non-targeted stocks. Bots also appear to be targeting stocks during important information events, as evidenced by the fact that on the day quarterly earnings are released they are double that of the rest of the sample. Not surprisingly, and as can be inferred in Table 1, attention is greater for bot-targeted stocks.

6.3 Bots as promoters: Impact on prices

What economic impact do bots have on a targeted stock? To examine this question, we define the variable of interest, *Bot promoted* as a dummy variable equal to one if a stock was mentioned on a given day by a bot, and the sentiment of the message was positive.²⁴

Insert Table 7

Table 7 reports the results from this analysis. The dependent variable is the market abnormal return. On the day bots are promoting a stock (day t), we find that the coefficient on *Bot promoted* is 1.19%, which implies that a stock that a bot promotes is associated with a higher 1.19% abnormal return compared to one that is not promoted. This is highly economically and statistically significant. Interestingly, there is a partial subsequent reversal in returns that persists up to 20 trading days following the promotion. For instance, the coefficient on *Bot promoted* in the last specification ($CAR_{t+1, t+20}$) is -0.38%. Thus, about one-third of the ‘pump’ from bots is subsequently reversed in the first 20 trading days after the promotion. It is important to note that we control for attention and sentiment from all other social media platforms and include firm and day fixed effects along with firm controls. The results support the hypothesis that bot activity causes substantial information distortions.

²⁴ We focus only on positive mentions for three reasons. They have a much larger incidence rate, retail investors have a strong preference for long positions (e.g., Bradley et al., 2024) and a bullish signal is much clearer than negative sentiment. We define a message as positive if the sentiment is greater than 0. Our results are robust and generally improve with a higher threshold; however, we chose 0 to avoid placing additional filters on our data and to maintain a higher incidence rate.

6.4 Bots as promoters: Order imbalances

In the previous section, we showed that when bots promote a stock, it leads to a significant increase in the stock price, followed by a subsequent reversal. In this section, we further evaluate how social media bots distort information by analyzing order imbalances. To measure retail volume, we follow Barber et al. (2023), who improved the algorithm by Boehmer et al. (2021). Specifically, we identify off-exchange trades in TAQ that have an exchange code of ‘D’ since many retail trades are filled off-exchange. We then compute the fractional-cent of the transaction price. Potential retail trades are transacted at prices between 0.0 and 0.4 (slightly below a penny) or between 0.6 and 1.0 (slightly above a penny). Half-penny trades are not considered because they may correspond to non-retail orders. Then, as in Lee and Ready (1991), we compare the transaction price to the midpoint of the National Best Bid and Offer. If the transaction price is higher than the midpoint, we sign it as a buy. Otherwise, it is a sell. We then follow Bradley et al. (2024) and focus on *Std. abnormal imbalance* as the retail order imbalance on day t , less the average imbalance over days $t-120$ through $t-240$, scaled by the standard deviation of imbalance estimated over the same 6-month estimation window. Our conclusion remains identical when using different definition of order imbalance. We report the results of our analysis in Table 8.

Insert Table 8

In model 1 of Table 8, we find that *Bot Promotion* is positive and statistically significant. The coefficient of 0.083 implies that when a stock is promoted by bots on *StockTwits*, retail order imbalance is 8.3% higher. With the exception of $t+4$, this order imbalance persists for the next five trading days. These results are similar to those of Bradley et al. (2024), who find that posts on *Wallstreetbets* can amplify trading frenzies among retail investors. Our results suggest that bots can have a similar effect

on retail trading frenzies.²⁵ We more explicitly test the impact bots have on trading frenzies in the next subsection.

6.5 Bots and trading frenzies

We follow Green and Jame (2025) who provide a definition of trading frenzies. We first calculate daily retail order imbalance as in section 6.3 and then take the average over the past 3 months (or 90 days). A stock is defined as having a retail buying frenzy if the calculated average order imbalance is over 2%.

In our data, retail trading frenzies are rare. We observe less than 1,000 trading frenzies from our sample of over 2.7 million observations. Given this low frequency, we reduce the sample to include only observations from the same industry where on the given day a trading frenzy occurred. This reduces our sample to around 140k observations. We estimate a linear probability model where *Trading frenzy* is a dummy variable that takes on a value of one if a buying trading frenzy occurred, zero otherwise. The independent variables are similar to Table 8. Industry and day fixed effects are included in the specifications as noted.²⁶ Results are reported in Table 9.

Insert Table 9

Model 1 of Table 9 shows the results with the most stringent fixed effects (industry and day). The coefficient *Bot promoted* is positive and statistically significant. The coefficient value of 0.003, or 0.3% is economically large given the sample mean is 0.68%. This suggests that bot promotion increases the likelihood of a retail buying frenzy by almost 50% relative to the sample mean. Frenzies are also

²⁵ While Bradley et al. (2024) did not examine bot activity on *Wallstreetbets*, it is certainly plausible that bots are active on the platform and amplify frenzies. In fact, it may be likely that the drop of informativeness of *Wallstreetbets* could have also been to large influx of bot accounts on the platform, following the rise in prominence after GME short selling squeeze.

²⁶ Similar to our analysis of PEAD in Table 4, we include industry fixed effects instead of firm fixed effects because trading frenzies are rare and firm fixed effects would likely absorb the vast majority of volatility due to over fitting the regression.

positively related to *Stocktwits* and *Dow Jones News* attention, *Twitter* sentiment, and volatility while negatively related to firm size and prior performance.

Given our previous results that indicate that following bot promotions, stock prices reverse when coupled with order imbalances, the evidence is consistent with retail traders exerting price pressure, possibly driving prices away from their fundamental level. In such cases, institutions trade against them, and retailers incur losses. Overall, these results are consistent with information distortions caused by bots and show how bots can be misused to initiate or amplify trading frenzies, leading to retail herding.

7. Conclusion

A voluminous literature focuses on analyzing social media to evaluate its impact on investment advice (e.g., Bradley et al., 2024), the behavior of retail investors, and their impact on capital markets (e.g., Cookson et al., 2024). It is widely acknowledged that bots are pervasive on social media platforms and can manipulate attention and opinion. Thus, it is possible that they also can have this effect on investment decisions. Yet, academic work has not considered this possibility. This is the focus of this study.

We first take advantage of shocks to CAPTCHA, a technology designed to deter bots from accessing media platforms, to explore the effects of bots. These technological shocks are exogenous to human activity on social media; the underlying time series variable is stationary, and we use day fixed effects in our specifications, which allows us to interpret the effect as causal.

We find that bot activity leads to higher price synchronicity and increases post-earnings announcement drift. These results suggest that bots increase information friction, distract investors from relevant information, and reduce informational efficiency. Our findings are consistent with the theoretical predictions in DellaVigna and Pollet (2009) and illustrate how bots can lead to information frictions, resulting in less informative prices.

We also take a different approach and identify accounts where we have a high degree of confidence that they are indeed bot accounts. In this setting, we can examine how bot promoters impact stock prices not only by distorting real attention but also through the content of their message, which could be incorrect or fabricated (Kogan et al., 2023). Consistent with the notion that bots distort prices, we find significant positive returns when bots promote a stock on day t , which subsequently reverse over the next 20 trading days. We further find that bots can significantly impact retail order imbalances, which can potentially amplify retail trading frenzies (Bradley et al., 2024).

Overall, our paper documents how bot activities can negatively impact social media discussions and the gathering of information. With the growing prevalence of AI and AI-driven bots, the risks of automated manipulation on social media are becoming increasingly severe (e.g., Dou et al., 2025). Given individuals' heavy dependence on social platforms for information (e.g., Cookson et al., 2024), our study shows how bots can be leveraged to facilitate market manipulation and undermine information efficiency.

References

- Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31-56.
- Ray Ball and Philip Brown, (1968). An Empirical Evaluation of Accounting Income Numbers, *Journal of Accounting Research*, 6(2), 159-178
- Barber, B. M., Huang, X., Jorion, P., Odean, T., & Schwarz, C. (2024). A (sub) penny for your thoughts: Tracking retail investor activity in TAQ. *The Journal of Finance*, 79(4), 2403-2427.
- Bartov, E., Faurel, L., & Mohanram, P. S. (2018). Can Twitter help predict firm-level earnings and stock returns?. *Accounting Review*, 93(3), 25-57.
- Ben-Rephael, A., Da, Z., & Israelsen, R. D. (2017). It depends on where you search: A comparison of institutional and retail attention. *Review of Financial Studies*, 30(9), 3009-3047.
- Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First Monday*, 21(11-7).
- Blankespoor, E., Miller, G. S., & White, H. D. (2014). The role of dissemination in market liquidity: Evidence from firms' use of Twitter™. *The accounting review*, 89(1), 79-112.
- Boehmer, E., Jones, C. M., Zhang, X., & Zhang, X. (2021). Tracking retail investor activity. *The Journal of Finance*, 76(5), 2249-2305.
- Bradley, D., Hanousek Jr, J., Jame, R., & Xiao, Z. (2024). Place your bets? The value of investment research on Reddit's Wallstreetbets. *The Review of Financial Studies*, 37(5), 1409-1459.
- Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., ... & Dredze, M. (2018). Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American Journal of Public Health*, 108(10), 1378-1384.
- Chakrabarti, P., Malvi, E., Bansal, S., & Kumar, N. (2023). Hashtag recommendation for enhancing the popularity of social media posts. *Social Network Analysis and Mining*, 13(1), 21.
- Chen, H., De, P., Hu, J., & Hwang, B. H. (2014). Wisdom of the crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, 27, 1367-1403.
- Chen, Q., Goldstein, I., & Jiang, W. (2007). Price informativeness and investment sensitivity to stock price. *The Review of Financial Studies*, 20(3), 619-650.
- Chen, X., He, W., Tao, L., & Yu, J. (2023). Attention and underreaction-related anomalies. *Management Science*, 69(1), 636-659.
- Clarke, J., Chen, H., Du, D., & Hu, Y. J. (2020). Fake news, investor attention, and market reaction. *Information Systems Research*, 32(1), 35-52.

- Cookson, J. A., Lu, R., Mullins, W., & Niessner, M. (2024). The social signal. *Journal of Financial Economics*, 158, 103870.
- DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and Friday earnings announcements. *The Journal of Finance*, 64(2), 709-749.
- Dou, W. W., Goldstein, I., & Ji, Y. (2025). Ai-powered trading, algorithmic collusion, and price efficiency. *Jacobs Levy Equity Management Center for Quantitative Financial Research Paper, The Wharton School Research Paper*.
- Duffy, C., Fung, B. (2022). Elon Musk commissioned this bot analysis in his fight with Twitter. Now it shows what he could face if he takes over the platform, CNN, available at <https://edition.cnn.com/2022/10/10/tech/elon-musk-twitter-bot-analysis-cyabra/index.html>
- Farrell, M., Green, T. C., Jame, R., & Markov, S. (2022). The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics*, 145(2), 616-641.
- Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday*, 22(8).
- Goldstein, I. (2023). Information in financial markets and its real effects. *Review of Finance*, 27(1), 1-32.
- Goldstein, I., Ozdenoren, E., & Yuan, K. (2013). Trading frenzies and their impact on real investment. *Journal of Financial Economics*, 109(2), 566-582.
- Goldstein, I., Xiong, Y., & Yang, L. (2025). Information sharing in financial markets. *Journal of Financial Economics*, 163, 103967.
- Green, T. C., & Jame, R. (2025). Retail Trading Frenzies and Real Investment. *Available at SSRN 4874993*.
- Gu, C., & Kurov, A. (2020). Informational role of social media: Evidence from Twitter sentiment. *Journal of Banking & Finance*, 121, 105969.
- Hanousek Jr, J., Hanousek, J., & Sokolov, K. (2025). X Bots and Earnings Announcements. *Available at SSRN 4595101*.
- Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance*, 64(5), 2289-2325.
- Hoffmann, M., Nagle, F., & Zhou, Y. (2024). The value of open source software. Harvard Business School Strategy Unit Working Paper, (24-038).
- Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225.

- Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political astroturfing on twitter: How to coordinate a disinformation campaign. *Political Communication*, 37(2), 256-280.
- Kelley, E. K., & Tetlock, P. C. (2017). Retail short selling and stock prices. *The Review of Financial Studies*, 30(3), 801-834.
- Kenny, R., Fischhoff, B., Davis, A., Carley, K. M., & Canfield, C. (2024). Duped by bots: why some are better than others at detecting fake social media personas. *Human factors*, 66(1), 88-102.
- Kogan, S., Moskowitz, T. J., & Niessner, M. (2023). Social media and financial news manipulation. *Review of Finance*, 27(4), 1229-1268.
- Lee, C., & Ready, M. (1991). Inferring investor behavior from intraday data. *The Journal of Finance*, 46, 733-746.
- Liu, H., Peng, L., & Tang, Y. (2023). Retail attention, institutional attention. *Journal of Financial and Quantitative Analysis*, 58(3), 1005-1038.
- Ouimet, P., & Tate, G. (2020). Learning from coworkers: Peer effects on individual investment decisions. *The Journal of Finance*, 75(1), 133-172.
- Wright, N. L., Nagle, F., & Greenstein, S. (2023). Open source software and global entrepreneurship. *Research Policy*, 52(9), 104846.
- Yang, K. C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., & Menczer, F. (2019). Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, 1(1), 48-61.

This figure is constructed by first selecting all *StockTwits* messages that mention a ticker and then sorted on those that occurred on days with the lowest quartile of bot activity from those that occurred on days with the highest quartile of bot activity. For each sample, we create a dictionary of words and calculate the frequencies after removing common English stop words (“I, me, ourselves, because ...”). We then subtract the frequencies of words in the top pentiles from their frequencies on days with low bot activity. The resulting data corresponds to the words most frequently mentioned on days of high bot access compared to days of low bot access.



“Short squeeze coming? Spy is moving fast again—could get crazy by the end of the day.”
“Can’t believe I missed that breakout. Price flew while I was busy reading old news.”
“Stocks going red again. Should I buy more or wait for the market to cool down?”
“Just sold my shares, not feeling the vibes right now. This market got no chill.”
“Feels like a good day to grab more. Price is low, time is right. Let’s go.”

Figure 2. Persistence of bot attention

In this figure, we show the persistence of attention on *StockTwits*. Specifically, we define attention as $\log(1 + \text{number of messages})$, and we separately calculate the pentiles of *StockTwits* attention on days with high and low bot access. Then, we calculate the means for the top pentile of both categories and analyze its persistence from day t until day $t+10$. The red line corresponds to attention on days with easier bot access, and the line corresponds to the y-axis on the right side. The figure is standardized to facilitate comparison on day t .

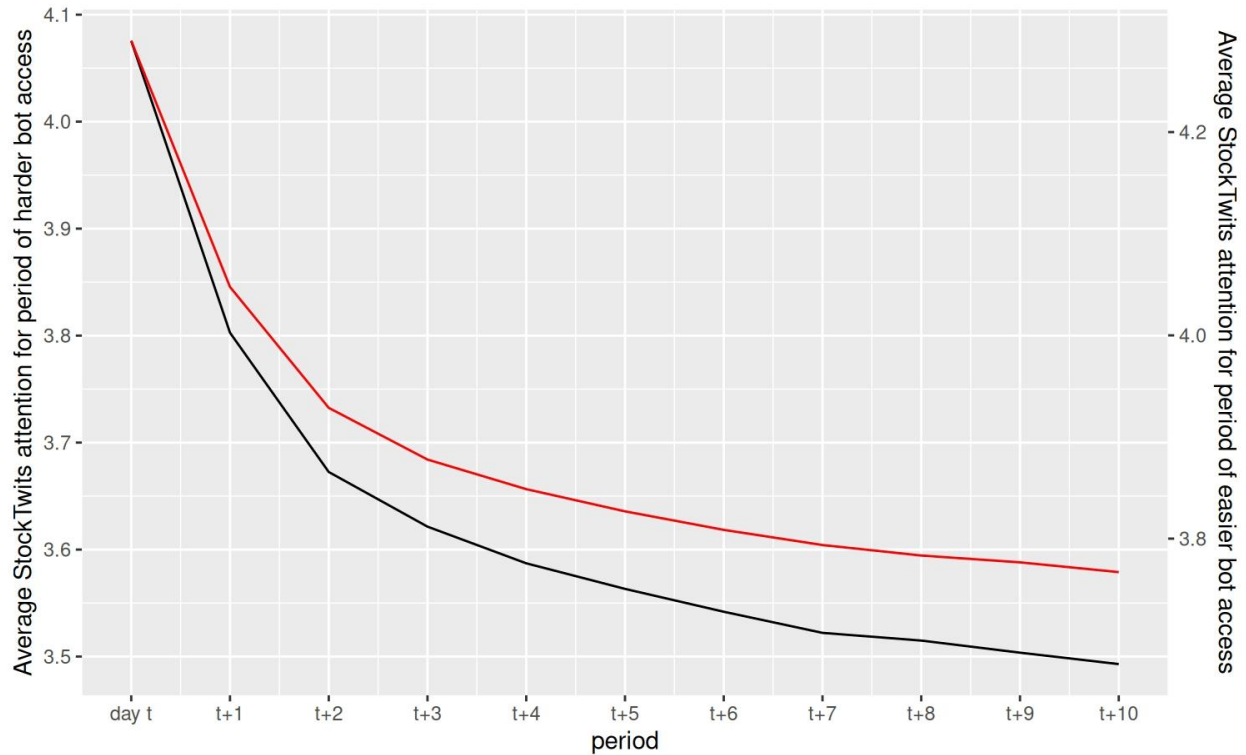


Figure 3. Distribution of attention

This figure shows the distribution of attention across days with easier (black) or harder (red) bot access on social media platforms. Attention is defined following Cookson et al. (2024): (number of messages) / (total number of messages on the day). We use $\log()$ transformation to achieve a smoother distribution.

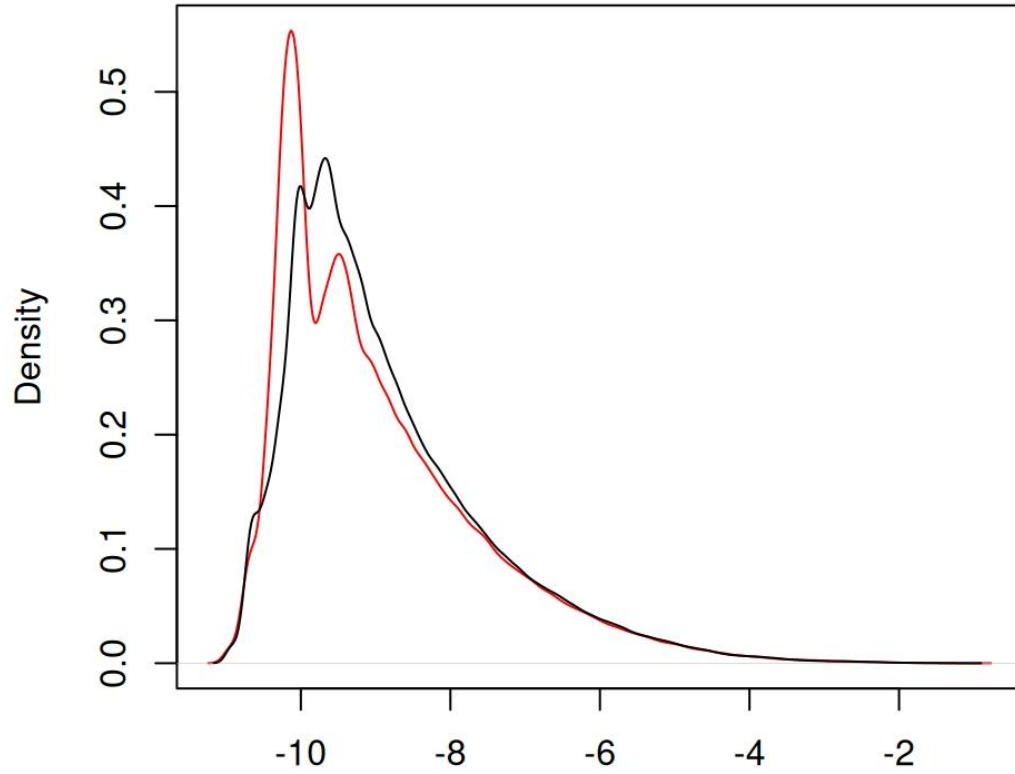


Figure 4. Post-earnings announcement drift and bot activity

The figure plots the effect of earnings announcements on day t to $t+40$ DGTW-adjusted returns for four cases: positive (negative) SUE with easier (harder) bot access. To estimate conditional returns, we calculate the group's SUE average (conditional mean) and multiply it by the group's relevant SUE regression coefficient as estimated in Table 4.

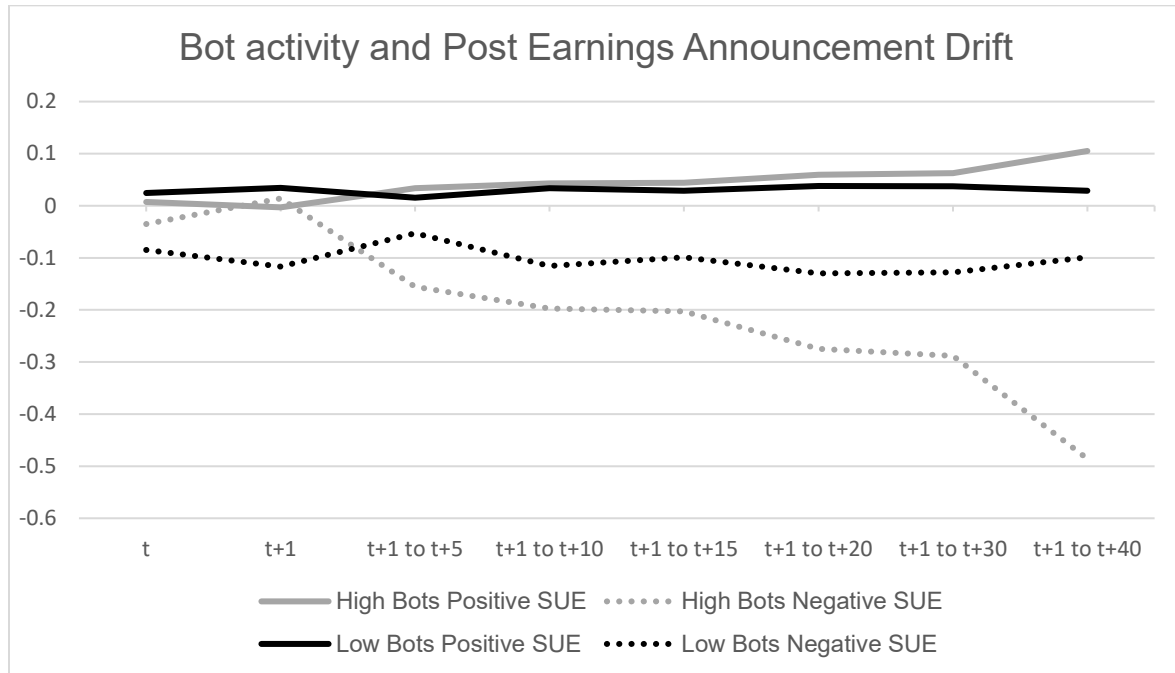


Table 1. Bots and social media attention

This table reports the impact of bot activity on attention on *Stocktwits* and *Twitter*. In Panel A, attention is defined as log (1 + the number of firm messages). We select the top 1,500 firms by message volume on *StockTwits* for each day since January 1, 2016. Our main independent variable is *CAPTCHA bugs*, which proxy for technological shocks to CAPTCHA technology, which is unresolved, but reported bugs for CAPTCHA programs in GitHub. In panel B, we define attention following Cookson et al. (2024) using PCA and across all platforms. For each platform, we report results using both raw and standardized measures of attention to ensure comparability with Cookson et al. (2024). We use their sample of the top 1,500 firms over their study period, but we restrict the sample to start in 2016 due to our proxies of bot activity. We drop observations with no messages on the platforms, consistent with Cookson et al. (2024). Firm and year fixed effects are included in all specifications. A full description of our methods is in Section 2 and Appendix A.

Panel A: Bots and modified attention

	StockTwits (log)	Twitter (log)
<i>CAPTCHA bugs</i>	0.3221*** (0.0035)	0.2749*** (0.0042)
Constant	1.0552*** (0.0143)	1.8150*** (0.0173)
Firm FE	Yes	Yes
R ²	0.524	0.483
R ² (adjusted)	0.522	0.482
N (obs)	2,744,374	2,184,037

Panel B: Bots and attention using PCA

	Cookson et al. (2024)	StockTwits		Twitter	
	1st PCA	Raw	Standardized	Raw	Standardized
<i>CAPTCHA bugs</i>	0.0146** (0.0059)	0.0003*** (0.0001)	0.0231*** (0.0074)	0.0003** (0.0001)	0.0264** (0.0112)
Constant	-0.0278 (0.0245)	0.0029*** (0.0004)	-0.0101 (0.0311)	0.0041*** (0.0005)	0.2027*** (0.0469)
Firm FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
R ²	0.371	0.324	0.324	0.440	0.440
R ² (adjusted)	0.370	0.322	0.322	0.438	0.438
N (obs)	683,494	603,505	603,505	494,982	494,982

Table 2. StockTwits' message characteristics during high and low bot access periods

This table presents summary statistics for various characteristics of *StockTwits* messages and platform activity. In Panel A, we show the characteristics during periods of low bot access, while Panel B shows the characteristics during periods of high bot access. The variables are either defined at the daily level, denoted as (daily), or at the message level, denoted as (message). We report the mean daily counts for the number of unique users, the number of messages, and the number of words per post. On *StockTwits*, users can flag their messages as bullish or bearish. We report the mean of the missing sentiment dummy variable across all messages. We also report how many messages were responses to other messages. For each variable, we report the number of observations, the standard deviation, the mean, and the 25th, 50th, and 75th percentiles.

Panel A: Low bot access periods

	N (obs.)	SD	mean	25%	50%	75%
N(users) (daily)	503	6,767	15,731	11,293	14,030	21,586
N(messages) (daily)	503	31,684	73,903	54,664	69,294	94,474
Words per post (daily)	503	2.17	17.67	15.90	16.71	19.73
Missing Sentiment (message)	37,173,262	0.49	0.59	0.00	1.00	1.00
Is a post a response to another post (message)	37,173,262	0.14	0.02	0.00	0.00	0.00

Panel B: High bot access periods

	N (obs.)	SD	mean	25%	50%	75%
N(users) (daily)	632	18,144	25,385	9,108	22,086	35,681
N(messages) (daily)	632	82,527	112,186	44,383	96,970	148,836
Words per post (daily)	632	2.17	18.60	16.65	19.20	20.03
Missing Sentiment (message)	70,902,009	0.50	0.51	0.00	1.00	1.00
Is a post a response to another post (message)	70,902,009	0.14	0.02	0.00	0.00	0.00

Table 3. Bot activity and informational efficiency

This table analyzes informational efficiency via price synchronicity during periods of high and low bot activity on social media. For each stock, we run the following regression model.

$$Ret_{i,t} = \alpha_i + \beta_{1,i}Market\ return + \beta_{2,i}Industry\ return$$

Price non-synchronicity is defined as $1 - R^2$, where lower values indicate better price informativeness, as prices and returns reflect more firm-specific information rather than industry or market trends. Consistent with the literature, we run the regression for each firm separately. We report the price synchronicity when employing the full sample. Then, we report price non-synchronicity, restricting observations to periods of easier and harder bot access.

	N (obs.)	SD	mean	25%	50%	75%
Price synchronicity (full sample)	9,099	0.276	0.749	0.632	0.858	0.957
– Harder bot access	8,431	0.306	0.726	0.607	0.851	0.954
– Easier bot access	8,124	0.309	0.655	0.445	0.752	0.917

Table 4. Bot activity and post-earnings announcement drift

This table analyzes the impact bot activity has on post-earnings announcement drift (*PEAD*). We follow Ben-Rephael et al. (2017) and Liu et al. (2023) in our empirical setup, focusing on quarterly earnings announcement days and the effect of standardized unexpected earnings (*SUE*) on future returns. We define *SUE* as the difference between actual EPS and analysts' median forecast, scaled by the share price at the time of the announcement. The dependent variable is the DGTW adjusted return over various intervals (day *t* representing the quarterly earnings release day). We use similar control variables to Cookson et al. (2024), but due to space constraints do not report *DJNW* attention and sentiment, indicators for QE and 8-K statement dates, size, volatility, number of analyst, and past returns. We report estimates during periods with harder (easier) bot access on social media platforms. Returns are reported as percentages. Standard errors are clustered at the firm and day levels.

	DGTW return (<i>t</i>)		DGTW returns (<i>t+1</i>)		DGTW returns (<i>t+1, t+10</i>)		DGTW returns (<i>t+1, t+20</i>)		DGTW returns (<i>t+1, t+40</i>)	
Variables	Harder bot access	Easier bot access	Harder bot access	Easier bot access	Harder bot access	Easier bot access	Harder bot access	Easier bot access	Harder bot access	Easier bot access
<i>SUE</i>	0.077** (0.033)	0.025 (0.034)	0.106*** (0.030)	-0.010 (0.092)	0.105 (0.148)	0.141* (0.072)	0.118 (0.171)	0.196** (0.083)	0.090 (0.192)	0.347*** (0.103)
<i>StockTwits attention</i>	0.050 (0.121)	0.014 (0.131)	-0.143 (0.109)	-0.479*** (0.109)	0.082 (0.214)	-0.555*** (0.195)	0.248 (0.220)	-0.781*** (0.241)	0.367 (0.345)	-0.147 (0.354)
<i>StockTwits sentiment</i>	1.961*** (0.439)	4.290*** (0.472)	-0.161 (0.618)	0.419 (0.645)	-0.505 (0.939)	1.729 (1.052)	-0.343 (1.131)	1.220 (1.317)	-0.573 (1.446)	4.098* (2.145)
<i>Twitter attention</i>	0.121* (0.068)	0.031 (0.087)	0.226*** (0.070)	0.115 (0.079)	0.167 (0.103)	0.283** (0.136)	0.245* (0.125)	0.333** (0.157)	0.441*** (0.150)	0.051 (0.227)
<i>Twitter sentiment</i>	2.931*** (0.348)	3.018*** (0.450)	0.039 (0.272)	-0.047 (0.310)	-0.432 (0.447)	-1.288*** (0.467)	-1.027* (0.527)	-0.729 (0.570)	-0.987 (0.658)	-1.045 (0.824)
<i>SeekingAlpha attention</i>	-0.350** (0.166)	-0.119 (0.188)	0.043 (0.160)	-0.164 (0.175)	0.367 (0.269)	-0.380 (0.314)	0.130 (0.349)	-0.062 (0.377)	-0.137 (0.446)	-0.156 (0.493)
<i>SeekingAlpha sentiment</i>	3.229*** (0.232)	3.078*** (0.326)	0.168 (0.118)	-0.194 (0.226)	0.028 (0.303)	0.163 (0.445)	0.066 (0.388)	-0.316 (0.583)	-0.076 (0.537)	-1.127 (0.708)
Constant	0.464 (1.194)	-1.716 (1.287)	1.827 (1.525)	-0.904 (1.538)	8.077*** (2.440)	-0.124 (2.467)	14.597*** (2.853)	4.664 (3.476)	20.222*** (4.020)	13.458*** (4.611)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.127	0.102	0.044	0.059	0.047	0.060	0.053	0.072	0.054	0.076
R ² (adjusted)	0.088	0.057	0.002	0.012	0.004	0.014	0.011	0.027	0.012	0.030
N (observations)	8,862	10,318	8,862	10,318	8,862	10,318	8,862	10,318	8,862	10,318

Table 5. CAPTCHA bugs and bot activity

This table analyzes the relationship between our measure of bot availability/supply: *CAPTCHA bugs* and the activity of bot accounts. *CAPTCHA bugs* corresponds to the natural logarithm of the number of issues (i.e., bugs) filed on GitHub for CAPTCHA providers. High values of CAPTCHA bugs indicate periods where CAPTCHA was not functioning properly, leading to easier platform access for bots. We identify bot accounts by 1) collecting accusations from other users that an account is a bot and, 2) manually verifying them. We consider the accusation valid if the account does not respond to the accusation but remains active. We then aggregate the number of messages the given bot account makes on the given day. The dependent variable is the number of messages per account and the natural logarithm of the number of messages, respectively. We use the bot account unique identifier as a fixed effect and cluster the standard errors at both date and bot account level.

	Number of messages	Log(Number of messages)
<i>CAPTCHA bugs</i>	10.61*** (1.98)	0.70*** (0.09)
Constant	-25.03*** (8.08)	-0.90** (0.38)
Bot FE	Yes	Yes
R ²	0.59	0.40
R ² (adjusted)	0.59	0.40
N (observations)	223,283	223,283

Table 6. Firms targeted by bots

This table contains comparison of firm characteristics between firms promoted by bots, i.e., firm-days where *Bot promoted* is equal to one, zero otherwise. We report the mean of variables when *Bot promoted* is equal to one and zero, and a *t*-test to test this difference. See Appendix A for a detailed description of variables.

	<i>Bot promoted</i>	Rest of the sample	Diff	t-stat
Size (log)	21.418	20.593	0.824***	114.160
Volatility	0.053	0.036	0.017***	64.801
Illiquidity	0.210	0.567	-0.357***	-5.536
CAR (t - 5, t - 1)	2.59%	0.38%	2.22%***	34.146
CAR (t - 30, t - 6)	3.35%	0.30%	3.05%***	29.368
StockTwits Attention (<i>log</i>) (t-3, t-1)	3.785	2.061	1.723***	376.600
StockTwits Attention (<i>log</i>) (t-7, t-1)	3.730	2.063	1.667***	372.020
StockTwits Attention (<i>ratio</i>) (t-3, t-1)	0.38%	0.05%	0.33%***	105.800
StockTwits Attention (<i>ratio</i>) (t-7, t-1)	0.37%	0.05%	0.32%***	107.180
QE Announcement date	0.052	0.026	0.027***	49.151

Table 6. Bot promotion and stock returns

This table analyzes the impacts of bot promotion on stock returns. We identify bot accounts by collecting accusations from other users that an account is a bot and manually verifying them. We consider the accusation valid if the account does not respond to the accusation but remains active. *Bot promoted* is a dummy variable equal to one if a bot account mentioned the stock ticker on day t , it was the only ticker in the message (consistent with Cookson et al., 2024), and the sentiment of the message was positive. The dependent variable is the abnormal return on days t , $t+1$, and the cumulative abnormal return on days $(t+1, t+2)$ to $(t+1, t+20)$. We use similar control variables to Cookson et al. (2024), but due to space constraints do not report *DJNW* attention and sentiment, indicators for QE and 8-K statement dates, size, volatility, number of analyst, and past returns. All return measures are calculated as percentages. Standard errors are clustered at the firm and day levels.

	AR(t)	AR($t+1$)	CAR($t+1, t+2$)	CAR($t+1, t+3$)	CAR($t+1, t+4$)	CAR($t+1, t+5$)	CAR($t+1, t+10$)	CAR($t+1, t+15$)	CAR($t+1, t+20$)
<i>Bot promoted</i>	1.19*** (0.05)	-0.01 (0.02)	-0.05 (0.03)	-0.09** (0.04)	-0.12** (0.05)	-0.14** (0.06)	-0.20** (0.09)	-0.33*** (0.12)	-0.38*** (0.15)
StockTwits attention	0.64*** (0.03)	-0.15*** (0.01)	-0.22*** (0.02)	-0.28*** (0.02)	-0.32*** (0.03)	-0.34*** (0.03)	-0.46*** (0.05)	-0.55*** (0.07)	-0.65*** (0.09)
StockTwits sentiment	1.06*** (0.03)	0.09*** (0.02)	0.09*** (0.03)	0.11*** (0.04)	0.15*** (0.04)	0.15*** (0.05)	0.16** (0.07)	0.12 (0.09)	0.07 (0.11)
Twitter attention	-0.29*** (0.02)	0.03*** (0.01)	0.04*** (0.01)	0.05*** (0.02)	0.05*** (0.02)	0.05** (0.02)	0.04 (0.04)	-0.01 (0.05)	-0.04 (0.06)
Twitter sentiment	1.29*** (0.04)	0.05* (0.02)	0.09** (0.04)	0.11** (0.04)	0.10** (0.05)	0.10* (0.06)	0.05 (0.08)	0.15 (0.10)	0.13 (0.12)
SeekingAlpha attention	-0.63*** (0.03)	0.03*** (0.01)	0.05*** (0.02)	0.08*** (0.02)	0.09*** (0.03)	0.12*** (0.03)	0.14*** (0.04)	0.16*** (0.05)	0.19*** (0.07)
SeekingAlpha sentiment	4.48*** (0.12)	-0.02 (0.03)	-0.09** (0.04)	-0.11** (0.05)	-0.14** (0.06)	-0.22*** (0.06)	-0.26*** (0.08)	-0.40*** (0.10)	-0.49*** (0.11)
Constant	-0.86*** (0.05)	0.23*** (0.03)	0.34*** (0.04)	0.42*** (0.06)	0.48*** (0.07)	0.53*** (0.08)	0.76*** (0.14)	0.95*** (0.20)	1.09*** (0.25)
Firm Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.04	0.03	0.04	0.04	0.05	0.05	0.06	0.07	0.09
R ² (adjusted)	0.04	0.02	0.03	0.04	0.04	0.05	0.06	0.07	0.08
N (observations)	2,722,401	2,720,311	2,718,331	2,716,599	2,714,681	2,712,747	2,709,961	2,707,341	2,704,808

Table 7. Social media attention and retail order imbalances

This table analyzes the impact of bot-promotion on retail order imbalances. We identify bot accounts by collecting accusations from other users that an account is a bot and manually verifying them. We consider the accusation valid if the account does not respond to the accusation but remains active. *Bot promoted* is a dummy variable defined as equal to one if a bot account mentioned the stock ticker on day t , it was the only ticker in the message (consistent with Cookson et al., 2024), and the sentiment of the message was positive. We define retail order imbalances using Bradley et al. (2024) as the retail order imbalance on day t , less the average imbalance over days $t-120$ through $t-240$, scaled by the standard deviation of imbalance estimated over the same 6-month estimation window. We analyze the impact from day t to $t+5$. We use the same set of controls as in Table 5, consistent with Cookson et al. (2024). Standard errors are clustered at the firm and day levels.

	Retail order imbalance					
	(t)	(t+1)	(t+2)	(t+3)	(t+4)	(t+5)
<i>Bot promoted</i>	0.083*** (0.006)	0.048*** (0.006)	0.022*** (0.006)	0.028*** (0.006)	0.013 (0.013)	0.029*** (0.006)
StockTwits attention	0.079*** (0.004)	0.044*** (0.003)	0.042*** (0.003)	0.042*** (0.004)	0.035*** (0.003)	0.031*** (0.003)
StockTwits sentiment	0.016*** (0.006)	0.021*** (0.006)	0.000 (0.006)	0.002 (0.006)	0.010* (0.006)	-0.003 (0.006)
Twitter attention	-0.015*** (0.003)	-0.007** (0.003)	-0.007*** (0.003)	-0.010*** (0.003)	-0.010*** (0.003)	-0.009*** (0.003)
Twitter sentiment	0.116*** (0.010)	0.070*** (0.009)	0.026*** (0.009)	0.025*** (0.009)	0.027*** (0.009)	0.024*** (0.009)
SeekingAlpha attention	-0.025*** (0.004)	-0.032*** (0.004)	-0.011*** (0.003)	-0.010*** (0.004)	-0.011*** (0.004)	-0.013*** (0.004)
SeekingAlpha sentiment	0.026*** (0.008)	0.050*** (0.008)	-0.025*** (0.007)	-0.011 (0.008)	-0.010 (0.008)	0.005 (0.008)
DJNW attention	-0.007*** (0.002)	-0.015*** (0.002)	-0.000 (0.002)	-0.004** (0.002)	0.001 (0.002)	0.001 (0.002)
DJNW sentiment	0.051*** (0.005)	0.038*** (0.005)	0.004 (0.005)	-0.000 (0.007)	-0.000 (0.005)	-0.012* (0.007)
8K form	0.011** (0.004)	-0.005 (0.004)	-0.008** (0.004)	-0.001 (0.004)	0.000 (0.004)	-0.007 (0.009)
QE day	0.140*** (0.008)	-0.031*** (0.007)	-0.062*** (0.006)	-0.011* (0.006)	-0.007 (0.006)	-0.026 (0.021)
Size	0.004 (0.005)	0.021*** (0.006)	0.018*** (0.006)	0.022*** (0.006)	0.024*** (0.006)	0.024*** (0.006)
Volatility	0.135*** (0.046)	0.586*** (0.041)	0.455*** (0.041)	0.406*** (0.042)	0.361*** (0.039)	0.289*** (0.039)
$CAR_{(t-5,t-1)}$	-0.317*** (0.018)	-0.461*** (0.017)	-0.380*** (0.016)	-0.347*** (0.016)	-0.315*** (0.015)	-0.268*** (0.014)
$CAR_{(t-30,t-6)}$	-0.099*** (0.008)	-0.090*** (0.008)	-0.078*** (0.008)	-0.070*** (0.008)	-0.063*** (0.008)	-0.057*** (0.008)
Constant	-0.192* (0.114)	-0.511*** (0.118)	-0.429*** (0.118)	-0.515*** (0.119)	-0.533*** (0.117)	-0.534*** (0.117)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.034	0.034	0.033	0.032	0.031	0.032
R ² (adjusted)	0.030	0.029	0.028	0.028	0.027	0.027
N (observations)	2,153,292	2,153,355	2,153,470	2,153,591	2,153,729	2,153,884

Table 8. Social media attention and trading frenzies

This table analyzes the impact of bot promotion on trading frenzies. We identify bot accounts by collecting accusations from other users that an account is a bot and manually verifying them. We consider the accusation valid if the account does not respond to the accusation but remains active. *Bot promoted* is a dummy variable defined as equal to one if a bot account mentioned the stock ticker on day t , it was the only ticker in the message (consistent with Cookson et al., 2024), and the sentiment of the message was positive. The dependent variable is a dummy variable for a trading frenzy defined following Green and Jame (2025) where we first calculate retail trading imbalance (in shares), scaled by total shares outstanding, and compute a 3-month (90-day) moving window. *Retail frenzies* equals 1 if the mean exceeds 2%. Since there are fewer than 1,000 retail frenzies in our sample, we restrict the sample to observations in which a retail frenzy occurred on the same day within the same Fama-French 48 industry. We use the same set of controls as in Table 5; however, due to the smaller sample and low incidence rate, we replace firm-fixed effects with Fama-French 48-industry fixed effects. Standard errors are clustered at the firm and day levels.

	Retail frenzies			
<i>Bot promoted</i>	0.003*	0.002	0.003**	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
StockTwits attention	0.005*	0.005*	0.005*	0.004*
	(0.003)	(0.002)	(0.003)	(0.002)
StockTwits sentiment	0.001	-0.000	-0.000	-0.001
	(0.004)	(0.004)	(0.004)	(0.004)
Twitter attention	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Twitter sentiment	0.007*	0.008**	0.006*	0.007*
	(0.003)	(0.004)	(0.003)	(0.004)
SeekingAlpha attention	-0.001	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
SeekingAlpha sentiment	0.001	0.001	0.000	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)
DJNW attention	0.001***	0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
DJNW sentiment	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
8K form	-0.000	-0.000	-0.000	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
QE day	-0.001	-0.001*	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Size	-0.004***	-0.004***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.000)	(0.000)
Volatility	0.156***	0.147***	0.155***	0.150***
	(0.048)	(0.044)	(0.048)	(0.046)
$CAR_{(t-5,t-1)}$	-0.045*	-0.042*	-0.045*	-0.042*
	(0.022)	(0.021)	(0.022)	(0.021)
$CAR_{(t-30,t-6)}$	0.012	0.011	0.012	0.012
	(0.009)	(0.009)	(0.009)	(0.009)
Constant	0.060***	0.063***	0.051***	0.049***
	(0.009)	(0.010)	(0.008)	(0.009)
Industry FE	Yes	Yes	No	No
Day FE	Yes	No	Yes	No
R ²	0.056	0.053	0.047	0.040
R ² (adjusted)	0.052	0.053	0.043	0.040
N (observations)	142,457	142,457	142,457	142,457

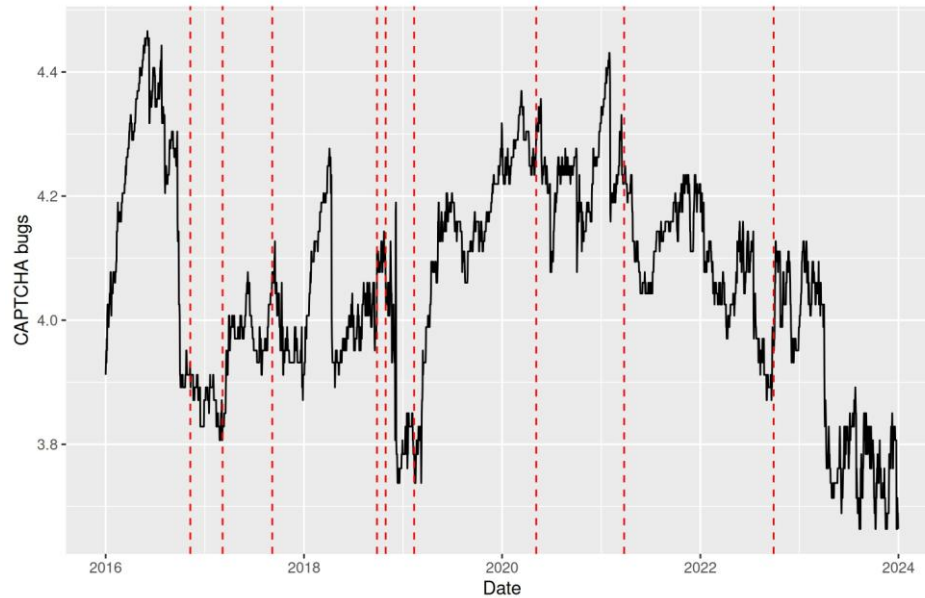
Appendix A. Shocks to bot activity

For our analysis of the effects of bot activity, we utilize technological shocks to enhance bot detection. One of the most common ways to limit bot activity is to employ the Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA). The application of CAPTCHA technology is relatively narrow, and the issues that disrupt it occur randomly. Therefore, shocks to CAPTCHA technology are unlikely to be related to investor behavior or preferences and unlikely to affect the company-specific financial variables through means other than bots.

Specifically, we collect unique data on innovations in CAPTCHA technology. Such innovations serve to curb bot activity. We rely on publicly available information on www.github.com to measure shocks to the CAPTCHA. GitHub is the largest developer platform for storing and managing open source code. Each application has a repository that includes documentation and all the necessary code and resources to run it. While our data relies on open-source resources, prior research (e.g., Hoffmann et al., 2024) shows that proprietary software developers often exploit them. We use textual analysis to categorize GitHub repositories as focusing on CAPTCHA implementation. We end up with 1,595 repositories in the CAPTCHA implementation. To estimate the dynamics of CAPTCHA implementation, we parse GitHub repositories to identify issues (i.e., bugs) reported and resolved in each repository. We then focus on the natural logarithm of the number of unsolved problems for CAPTCHA implementation repositories, which we refer to as *CAPTCHA bugs*. A larger number of *CAPTCHA bugs* should correspond to periods where there were likely many different exploits for bots to infiltrate the platform. Conversely, a low number of *CAPTCHA bugs* should indicate periods when the system operates well, and it is more difficult for bots to access the platform.

We plot the CAPTCHA bugs across our sample period, i.e., from January 2016 to December 2023, in the figure below. The red vertical dotted lines correspond to major shocks to CAPTCHA

technology, which we list below the figure. Let us note that the KPSS test does not reject the null hypothesis that the series is stationary.



2022-09-28	introducing CAPTCHA alternative called Turnstile (Cloudflare)
2021-03-26	introducing Super Bot Fight Mode (Cloudflare)
2020-05-06	Cloudflare Bot Management: machine learning and more (Cloudflare)
2019-02-11	Admin Console v3 Launch. (reCaptcha)
2018-10-29	reCAPTCHA v3 Launch. (reCaptcha)
2018-09-27	Strict style-src support added. (reCaptcha)
2017-06-09	reCAPTCHA Android Library Launch. (reCaptcha)
2017-03-07	Invisible reCAPTCHA Launch (reCaptcha)
2016-08-11	CSP nonce and 'strict-dynamic' support added . (reCaptcha)

In our paper, we oftentimes use different specifications and instead focus on periods of easy and harder bot access. We define those periods as those with bot activity in the bottom 25% or top 25% pentiles. Note that using smaller quantiles (10% or 20%) yields similar results. We use this specification because it allows us to more clearly identify and interpret the effect, while enabling the application of day FE, which can control for a variety of other effects.