

Choudhury, Prithwiraj; Vanneste, Bart S.; Zohrehvand, Amirhossein

Working Paper

The wade test: Generative AI and a CEO bot

CESifo Working Paper, No. 11316

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Choudhury, Prithwiraj; Vanneste, Bart S.; Zohrehvand, Amirhossein (2025) : The wade test: Generative AI and a CEO bot, CESifo Working Paper, No. 11316, Munich Society for the Promotion of Economic Research - CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/338288>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

CES ifo

**11316
2026**

Working Papers

Original Version: August 2024
This Version: March 2026

**The Wade Test: Generative AI
and a CEO Bot**

Prithwiraj Choudhury, Bart S. Vanneste, Amirhossein Zohrehvand

CES ifo

Imprint:

CESifo Working Papers

ISSN 2364-1428 (digital)

Publisher and distributor: Munich Society for the Promotion
of Economic Research - CESifo GmbH

Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740

Email office@cesifo.de
<https://www.cesifo.org>

Editor: Clemens Fuest

An electronic version of the paper may be downloaded free of charge

- from the CESifo website: www.ifo.de/en/cesifo/publications/cesifo-working-papers
- from the SSRN website: www.ssrn.com/index.cfm/en/cesifo/
- from the RePEc website: <https://ideas.repec.org/s/ces/ceswps.html>

The Wade Test:
Generative AI and a CEO Bot

Prithwiraj (Raj) Choudhury

Management Department
London School of Economics and Political Science
p.choudhury4@lse.ac.uk

Bart S. Vanneste

UCL School of Management
University College London
b.vanneste@ucl.ac.uk

Amirhossein Zohrehvand

Faculty of Science, Science Based Business
Leiden University
a.h.zohrehvand@sbb.leidenuniv.nl

February 27, 2026

Acknowledgement: We thank Xi Kang for the discussions that motivated this project. We are grateful for the feedback of Isabel Fernandez-Mateo, Blaine Landis, and Jen Rhymer.

ABSTRACT

Generative AI (Gen-AI) is increasingly used to communicate on behalf of organizational leaders. We investigate whether Gen-AI can mimic a CEO and whether employees display aversion to its communication. We present a framework distinguishing between a strong or weak form of Gen-AI aversion. In a field experiment, employees evaluated responses from their CEO (“Wade”) and an AI bot trained on the CEO’s communication. Employees correctly identified the source 59% of the time, yet rated responses believed were AI-generated as less helpful, regardless of actual source. A second study with a general audience directly manipulates AI labels and likewise finds AI-labeled responses are rated as less helpful. Together, the findings provide evidence of strong algorithm aversion to Gen-AI used as a CEO bot.

MANAGERIAL SUMMARY

Can an AI bot convincingly mimic a CEO’s? Extending the Turing test, we tested this with employees of a firm using an AI bot that mimicked its CEO. Employees correctly identified the response source only 59% of the time, suggesting the AI closely mimicked the CEO. However, employees rated messages they believed came from AI as less helpful, regardless of their actual source. This aversion was strongest for communication aimed at revealing aspects of the leader. A second study with a general audience confirmed these findings. CEOs considering AI for their communication should recognize that while AI can mimic them effectively, employees may resist a CEO bot, particularly for communication that reveals aspects of the CEO.

1. INTRODUCTION

How artificial intelligence (AI) will affect firms is a question of great interest to both organizational scholars (e.g., Allen & Choudhury, 2022; Berg et al., 2023; Doshi et al., 2025; Eloundou et al., 2024; Felten et al., 2021; Gaessler & Piezunka, 2023; Jia et al., 2024; Otis et al., 2024) and practitioners (e.g., Microsoft, 2024). Generative AI (Gen-AI) models, particularly large language models (LLMs), have emerged as transformative due to their ability to generate natural language, the primary form of communication in organizations (e.g., Yates & Orlikowski, 1992). Accordingly, research has investigated how Gen-AI is transforming the nature of work for employees (e.g., Brynjolfsson et al., 2025; Dell'Acqua et al., 2026; Noy & Zhang, 2023). Increasingly, employees are expected to interact and communicate, not only with other humans but also with AI bots. These can be designed to mimic specific individuals, possess knowledge about them, and even represent them in meetings (Leong et al., 2024). For example, some organizations, such as Zapier and Otter, have created AI bots to impersonate their CEOs, enabling interactions between employees and a CEO bot (Boyle, 2025).

An important caution to using personalized AI bots to mimic organizational communication is raised by the literature on algorithm aversion (e.g., Dietvorst et al., 2015; Jussupow et al., 2024). Algorithm aversion refers to people's tendency to discount or reject algorithmic output in favor of human output (e.g., Burton et al., 2020; Jussupow et al., 2024). Synthesizing prior literature, we can distinguish between a weak and strong form of algorithm aversion. While the weak form of algorithm aversion describes a discounting of algorithmic output when its performance is seen as worse than human output (e.g., Berger et al., 2021; Dietvorst et al., 2018), the strong form of algorithm aversion depicts a discounting of algorithmic output regardless of its actual or perceived performance, even when algorithms outperform humans (e.g., Jussupow et al., 2024; Yeomans et al., 2019). Hence, while a CEO bot imitating a specific CEO might be technologically feasible, whether such communication is well received by organizational actors remains a question of vital interest to organizational scholars. Therefore, we ask:

Can Gen-AI convincingly impersonate a CEO in their communication, and do recipients display aversion to Gen-AI communication?

We assess our research question in a field experiment with a firm and its CEO, where the firm took the lead in developing an AI bot intended to impersonate the CEO and automate asynchronous CEO communication. The experiment is inspired by the Turing test, originally known as the imitation game (Turing, 1950), which is a blind test where a human guesses whether they are communicating with a human or a machine. We extend the Turing test from imitating a *generic* person to imitating a *specific* person, namely the CEO, to learn if an employee can guess whether they are communicating with the CEO or the CEO bot that impersonates the CEO. We call this test the “Wade test,” named after the firm’s CEO. A total of 105 employees of the firm voluntarily participated in the Wade test, each responding to a set of 10 predefined questions and answers. Experimental results show that employees correctly guess the source’s identity in 59.0% of 1,050 instances, somewhat better than random chance. For six out of 10 questions, accuracy is not statistically different from 50%, suggesting that employees cannot distinguish between the bot’s answers and those of the CEO. In sum, the CEO bot nearly passes the Wade test. Next, we present evidence related to the second part of our research question: whether recipients exhibit algorithm aversion towards the Gen-AI communication. Employees rate answers they believe (often inaccurately) to be generated by the machine as less helpful than those they believe to be generated by the human CEO, which indicates Gen-AI aversion. Furthermore, we find that Gen-AI aversion is greater as Gen-AI performs better (i.e., when detectability is lower), which suggests a strong form of algorithm aversion. Relatedly, aversion is even more salient when the communication concerns revealing aspects of the CEO.

We conduct an additional experiment with a general audience to directly test for a strong form of Gen-AI aversion. Participants were presented with analysts’ questions posed during earning calls to the CEOs of Hershey, General Motors, Nvidia, and Amazon and the answers of either the CEO or an AI (an LLM instructed to mimic the CEO in both content and style). We randomized whether answers were labeled as being generated by the human CEO or by the CEO bot. Importantly, regardless of their actual origin, answers labeled as AI generated were seen as less helpful than those labeled as generated by the CEO. In sum, whether the message is perceived by receivers as generated by AI (study 1) or labeled by

researchers as being generated by AI (study 2), receivers report lower helpfulness (i.e., greater algorithm aversion). These findings indicate that people exhibit a strong form of Gen-AI aversion in the context of a bot impersonating a CEO.

This study contributes to two research streams. First, we contribute to the burgeoning organizational literature on the impact of Gen-AI on firms. Whereas this literature has focused on performance of employees while using Gen-AI based tools, our study sheds light on the promise and challenges of developing and deploying personalized bots that mimic communication by a specific organizational actor, in our case the CEO. The experimental results suggest that the successful deployment of personalized bots needs to account for a strong form of algorithm aversion that human recipients might exhibit towards such communication, especially if such communication is revelatory of the CEO. Second, we contribute to the algorithm aversion literature by developing a conceptual framework of Gen-AI aversion that distinguishes between a weak and strong form of aversion and summarizes three distinct types of Gen-AI failures: hallucination, anthropomorphism failure, and impersonation failure.

2. CONCEPTUAL FRAMEWORK

Prior literature on algorithm aversion has predominantly focused on predictive algorithms (Jussupow et al., 2024). Predictive algorithms include supervised learning models and recommender systems. In the former, the goal is to predict an outcome (e.g., estimating a movie's box-office returns from features like genre, budget, and cast). In the latter, the goal is to predict the item that will maximize a user's satisfaction or engagement (e.g., recommending a movie to a user based on past watch history and ratings). Both types of algorithms learn from historical data to uncover patterns that link inputs to outputs, and then apply the learned mapping to new inputs to predict outputs. Gen-AI also learns patterns from historical data, but it shifts the goal from prediction to the creation of novel data (e.g., producing a new movie).

Considering this difference, we argue that prior literature’s insightful findings need to be reviewed and substantially extended to the case of Gen-AI applications, such as a CEO bot. To this end, we develop a conceptual framework of Gen-AI aversion (see Figure 1). Specifically, we take three steps. First, we distinguish between weak and strong forms of algorithm aversion. Second, we discuss unique sources of failures for Gen-AI beyond ways in which predictive algorithms could fail.¹ Finally, we link our Gen-AI aversion conceptual framework to the purpose of communication generated by a CEO bot.

[[INSERT FIGURE 1 ABOUT HERE]]

2.1. Weak and Strong Forms of Algorithm Aversion

Extant literature defines algorithm aversion as the tendency of people to systematically discount or reject algorithmic output in favor of human output (e.g., Burton et al., 2020; Dietvorst et al., 2015; Jussupow et al., 2024). For example, Dietvorst et al. (2015) find that, even when an evidence-based algorithm outperforms a human judge at predicting MBA students’ future success, participants still discount the algorithm’s forecasts and prefer the human’s instead. Many other studies also report evidence of algorithm aversion (e.g., Allen & Choudhury, 2022; Castelo et al., 2019; Dietvorst et al., 2018; Tong et al., 2021, but see Logg et al., 2019).

Synthesizing this literature, we distinguish between a weak and strong form of algorithm aversion. *Weak algorithm aversion* arises when people discount algorithmic output when its performance is perceived as worse than human output (e.g., Berger et al., 2021; Dietvorst et al., 2018). For example, in an incentivized call-center forecasting task, participants who observed the algorithm make small errors judged it less reliable and shifted to the human forecaster, discounting the algorithm because they perceived it would perform worse (Berger et al., 2021). In the case of predictive algorithms, performance is typically measured in terms of lower prediction errors: fewer misclassifications in a classification task (e.g., diagnosing a medical condition), predicted values closer to actual values in a regression task (e.g., estimating energy usage), or suggesting more preferred content in a recommendation task (e.g.,

¹ We refer to Gen-AI failures as they apply not only to LLMs but also SLMs (small language models), an increasingly popular form of Gen-AI (e.g., Schick & Schütze, 2021), and to multiple data modalities (e.g., text, speech, video).

suggesting a song on a streaming platform). The algorithm need not perform objectively worse; rather, it is sufficient if it is believed to perform worse (e.g., Dargnies et al., 2026). Weak algorithm aversion can be mitigated by demonstrating superior algorithmic performance or the algorithm's ability to learn (Berger et al., 2021; Dargnies et al., 2026).

Strong algorithm aversion, on the other hand, is a discounting of algorithmic output regardless of its performance, even when algorithms actual or perceived performance exceeds that of humans (Dietvorst et al., 2015; Önkal et al., 2009). For example, Longoni et al. (2019) find that when patients must decide whether to undergo coronary-bypass surgery, they follow a physician's prediction on the best treatment far more often than a computer program's, even though both sources are described as achieving identical diagnostic accuracy in interpreting the patient's blood-pressure, ECG, cholesterol and artery-blockage data. Past research documents other instances where people perceive algorithms as unfit to perform tasks, such as predicting how funny a joke is (Yeomans et al., 2019) or deciding whether to grant an offender parole based on the expected reoffending risk (Bigman & Gray, 2018). People may believe that algorithms categorically lack qualities that are required to perform certain tasks (e.g., prediction tasks with a higher degree of subjectivity) (Castelo et al., 2019).

2.2. Gen-AI Aversion and Sources of Gen-AI Failures

Gen-AI differs from predictive algorithms in two notable aspects: the scope of tasks performed and the types of algorithmic failures. Whereas predictive algorithms analyze existing data to predict outcomes, Gen-AI produces novel data. This remarkable capability to generate content gives rise to fundamentally different tasks, including code generation (Chen et al., 2021), scientific knowledge production (Bubeck et al., 2023), creative composition (Chen & Chan, 2024; Doshi & Hauser, 2024), and conversational interaction (Ross et al., 2023).

This shift toward open-ended generation leads to new sources of algorithmic failure. Whereas predictive algorithms have one primary source of algorithmic failure (i.e., inaccurate prediction), our conceptual framework outlines three sources of algorithmic failures relevant to Gen-AI aversion in a CEO bot context: hallucination, anthropomorphism failure, and impersonation failure.

Hallucination. A hallucination occurs when a Gen-AI model fabricates plausible but incorrect content (Kaddour et al., 2023). For example, a CEO bot might incorrectly claim that subscription-based business models are rapidly gaining traction in the company’s core market based on trends in a related industry. Although related to errors made by predictive algorithms, hallucinations are context-dependent and harder to identify systematically than those of predictive algorithms, whose prediction errors can be verified against a known ground truth (Ji et al., 2023). Because LLMs are built to generate snippets of text sequentially and probabilistically, they produce text that is both coherent and unpredictable, resulting in hallucinations (Zhang et al., 2023).

Anthropomorphism failure. Anthropomorphism failure is specific to Gen-AI designed to simulate human-like content (e.g., a customer service bot is expected to generate human-like text, unlike a code generator). Anthropomorphism failure occurs when Gen-AI falls short of natural human behavior in noticeable ways. For example, despite their sophisticated language capabilities, LLMs may display unnatural patterns such as being overly formal, maintaining an unnaturally consistent emotional tone, or failing to pick up on subtle social cues (Bojić et al., 2025). They might generate content with factually correct but socially awkward responses, such as being excessively polite, providing unnecessarily detailed explanations, or maintaining an unnatural level of patience that betrays their non-human nature (Kim et al., 2024). Unlike hallucination, where the model invents nonexistent facts, these errors arise despite factual accuracy as the response violates ordinary social expectations (Kim et al., 2024), highlighting anthropomorphism as a distinctive form of failure. It is these human-like characteristics that have long contributed to the challenge of passing the Turing test.

Impersonation failure. Impersonation failure, unique to Gen-AI built to impersonate a particular person, arises when the generated content misrepresents the focal human’s identity, opinions, or communication style. It can take several forms, including conveying personal experiences that happened but did not involve the individual, deviating from the individual’s established persona (e.g., extraversion), failing to adopt the individual’s communication style, or shifting communication styles inconsistently across conversations (Frisch & Giulianelli, 2024). Impersonation failure is therefore distinct from the two

other failure types. Whereas hallucination concerns truth about the general state of the world, and anthropomorphism failure concerns conformity to generic human social norms, impersonation failure concerns person-specific coherence: the generated text may be both factually correct and socially plausible yet still defective because it attributes those facts or behaviors to the wrong individual or in the wrong voice.

Although hallucination, anthropomorphism, and impersonation failures are conceptually distinct, they are interrelated. Their interdependence is most evident in the link between anthropomorphism and impersonation failures: if Gen-AI fails to appear convincingly human, it cannot reliably portray a specific human. Other failure modes may also interact. Paradoxically, injecting occasional hallucinations may enhance perceived humanness by mirroring the imperfect nature of human speech (Bluvstein et al., 2024) or by giving the impression of agency (Vanneste & Puranam, 2024). At the same time, hallucinations can undermine the impersonation of a person with integrity and help with establishing the impersonation of a person with low integrity. Just as the Turing test (i.e., imitating a generic person) depends on what is said and how it is said, the Wade test (i.e., imitating a specific person) is affected by the occurrence of all three failures.

2.3. Gen-AI Aversion in the Context of a CEO Bot

Prior research highlights communication as a critical task performed by CEOs (Bandiera et al., 2020; Eklund et al., 2026; Hambrick, 2007; Kotter, 1990). CEOs communicate with stakeholders within and across organizations to serve a wide range of purposes such as informing about decisions, affecting employees, and revealing personal facts (e.g., Kaplan, 2008; Liu et al., 2023). Communication is a time-consuming task for CEOs, and Bandiera et al. (2020) document that CEOs spend on average 70% of their time communicating with others.² Porter and Nohria (2018: 5) draw attention to the issue of CEOs being overwhelmed with communication in the digital age, and call on scholars to study how CEO communication can “stay ahead of the digital avalanche”. Gen-AI enables generic chatbots capable of

² Relatedly, Porter and Nohria (2018) report that CEOs spend about 40% of their time in non-face-to-face communication activities, with on average 15% of their time allocated to reading and replying to written correspondence and 24% to electronic communication.

communicating similarly to humans. Personalized bots, such as a “CEO bot”, go beyond generic chatbots by impersonating an individual and may have profound implications for how CEOs communicate with other stakeholders.

CEO bots are Gen-AI chatbots designed to mimic a specific executive and interact with employees and other stakeholders. These bots can facilitate CEO communication by providing an online channel allowing for synchronous or asynchronous communication, complementing direct communication from the CEO. As Gen-AI capabilities advance, particularly in language modeling and personalization, it is becoming increasingly feasible to deploy CEO bots that deliver contextually relevant communication at scale that is closely aligned with the CEO’s style. However, Gen-AI aversion introduces a critical concern for integrating a CEO bot into CEO communication: even if Gen-AI communication closely mimics that of the CEO, the audience may perceive and evaluate it differently.

In our investigations, we explore three steps: whether Gen-AI aversion exists in the context of CEO communication, whether any aversion is of the weak or strong form, and if Gen-AI aversion depends on the purpose of CEO communication. Detecting weak or strong forms of aversion toward CEO bot communication requires observing whether the aversion is less when the bot performs better (weak form of aversion) or whether the aversion is greater or unchanged when the bot performs better (strong form of aversion). Mirroring the Turing test, an important measure of the CEO bot’s performance relates to detectability. A higher-performing bot will make it difficult for message recipients to distinguish between bot-generated and genuine CEO communications, leading to lower detection rates. Given this, we posit the following: if we observe less Gen-AI aversion when the bot performs better (i.e., has lower detectability levels), that would be indicative of weak aversion toward the communication generated by the bot. In contrast, if Gen-AI aversion is greater or unchanged when the bot performs better, that would be indicative of strong aversion toward bot-generated communication.

Further, we contend that the degree of Gen-AI aversion toward CEO bot communication depends on the purpose of such communication. Through a comprehensive review of the leader communication literature, Liu et al. (2023) identify three main purposes of CEO communication: (1) to inform about

leadership decisions, (2) to affect and motivate stakeholders, and (3) to reveal aspects of leaders themselves. Gen-AI failures may not be equally relevant across these three communication purposes. For example, when the goal is to inform about leadership decisions, factual information (i.e., avoiding hallucinations) matters, whereas humanness (anthropomorphism failure) and personalness (impersonation failure) may matter less. Likewise, when the aim is to motivate stakeholders, factualness and humanness of the communication matter, whereas impersonation may matter less. However, if the purpose is to reveal aspects of the leader, all three failures could be relevant. Hence, below we focus on communication that reveals aspects of the CEO.

When a CEO bot reveals aspects of the CEO, the effect on Gen-AI aversion depends on whether that aversion is of the strong or weak form. We speculate that revelatory communication may amplify Gen-AI aversion if it is of the strong form. Revelatory communication may include personal facts, personality traits, or inner psychological processes (Akstinaite et al., 2020; Liu et al., 2023). These disclosures enhance perceptions of a leader's authenticity (Jiang et al., 2022), namely that a leader is genuine and acts in line with their true selves (Cha et al., 2019; George et al., 2007). Authentic leadership has received increased attention, because of the benefits to employees and organizations it yields (Clapp-Smith et al., 2009; Lehman et al., 2019; Rego et al., 2015; Wang & Hsieh, 2013; Wang et al., 2014). Because authentic leadership is based on genuineness and self-alignment, we anticipate that a CEO bot will incur a penalty when engaging in revelatory communication, even if it perfectly mimics the CEO (i.e., the strong form of Gen-AI aversion). Stakeholders may prefer personal disclosures straight from the CEO rather than via a CEO bot.

In contrast, revelatory communication may attenuate Gen-AI aversion if it is of the weak form. Because these disclosures concern private and difficult-to-observe experiences, recipients will have difficulty verifying the communication and thus struggle to differentiate bot-generated from genuine CEO messages. For example, recipients may find it easier to spot incorrect statements by a CEO bot if the question is about the current competitive landscape than if it is about a personal lesson learned by the CEO. Because a CEO bot's performance is inversely related to detectability, a CEO bot may

paradoxically perform better in revelatory communication. If so, then it may lessen the weak form of Gen-AI aversion, as it is performance contingent.

3. STUDY 1

We conducted a pre-registered field experiment with the company Pulse (a pseudonym) to assess (1) the extent to which employees can distinguish between CEO and AI messages and (2) whether a message's perceived origin affects its perceived helpfulness.³

3.1. Context

Pulse is a technology company with more than 800 employees. The CEO co-founded the company and is well-known internally. The CEO was motivated to engage in this study for two reasons. First, the Pulse CEO, like other CEOs, is shadowed by executive assistants who often respond to communications on the CEO's behalf. The CEO was intrigued by the prospect of automating this process. Second, leading a technology company, the CEO wanted to see whether the CEO bot used in this study could serve as a test case for future products.

The company led the development of the CEO bot, which consists of two LLMs. The underlying principle was that for the bot to be effective, it would have to mimic both the content and the style of CEO communication. The first model focused on content. It involved a foundation model (e.g., GPT-4) that answered questions based on company documents (e.g., Slack data, emails, and company reports). These documents were stored in a database, which the model used via retrieval augmented generation (RAG) (Gao et al., 2023). This technique integrates real-time document retrieval, reducing hallucination and improving factual correctness. The second model focused on style. Its goal was to rephrase input to impersonate the CEO. It involved a foundation model that was fine-tuned using texts the CEO had written (such as emails, Slack messages, and blog posts) addressed to both internal and external audiences. Thus, the user's question was the input to the first model. Its output (i.e., a response intended to be factually

³ The pre-registration is available here: https://aspredicted.org/NB1_F42

correct) served as input to the second model. The output of the second model (i.e., the response in Wade's style) was returned to the user.

3.2. Methods

3.2.1. Participants

All Pulse employees were invited to participate in the study on a voluntary basis (i.e., without extra compensation and beyond their usual job duties), and 105 employees (approximately 13 %) participated. Participants included senior management (13%) and other employees (87%). Participants had worked at the company for less than one year (13%), one to less than three years (44%), or three years or more (43%). Hence, the typical participant had been exposed to the CEO's communication over a sustained period. A comparison of seniority and tenure found that senior management is overrepresented in the sample compared with the organization overall (13.0% vs. 6.0%) and no statistical differences for tenure (see Table A11 in the Online Appendix). In a robustness check, we investigate how seniority and tenure affect the results.

3.2.2. Procedure

Similar to a Turing test (Turing, 1950), employees had to indicate whether they thought an answer to a question came from the CEO or the AI. The questions were selected from a pool of 148 questions that new hires had recently submitted in an Ask-Me-Anything (AMA) forum. From this pool, an LLM (distinct from the models powering the CEO bot) extracted 10 themes and one exemplar question per theme. We used the 10 resulting questions (see Table 1).⁴ The CEO and AI separately responded to the questions. All answers were kept unchanged.

[[INSERT TABLE 1 ABOUT HERE]]

⁴ The typical case is that focal respondents assess questions they did not ask. Out of a total of 1,050 observations, at most 10 will be for questions posed by a focal respondent.

In a within-subjects design, participants were presented with all 10 questions in the same order. Each employee randomly received the CEO’s answer for about half of the questions and the AI’s answer for the other questions.⁵ This approach yielded 105 (employees) * 10 (questions) = 1,050 observations.

3.2.3. Measures

Perceived origin was measured by the item “Who do you think wrote this answer?” (CEO / AI) and perceived helpfulness by the item “How helpful is this answer?” (1 = not helpful at all, 5 = extremely helpful).⁶ Because algorithm aversion is a tendency to discount or reject algorithmic output in favor of human output (e.g., Dietvorst et al., 2015; Jussupow et al., 2024), a difference in perceived helpfulness between responses believed to come from AI and those from the CEO is an indicator of Gen-AI aversion (Simantov-Nachlieli, 2025; Zhang et al., 2024).

3.3. Results

The accuracy of perceived origin was 59.0% (95% confidence interval (CI): [0.556, 0.624]). Guessing randomly would yield an approximate accuracy of 50%. Hence, distinguishing between CEO and AI answers is challenging for employees. Table 2 provides the confusion matrix, indicating that employees correctly identified 61.1% of CEO answers and 56.6% of AI answers. Figure 2 shows the accuracies and 95% CIs per question. Accuracies for six out of 10 questions are statistically indistinguishable from 50%.

[[INSERT TABLE 2 ABOUT HERE]]

[[INSERT FIGURE 2 ABOUT HERE]]

Model 1 of Table 3 reports a logistic regression with perceived AI (i.e., an indicator variable that is 1 when AI is the perceived origin and 0 otherwise) as the dependent variable. The specification includes fixed effects for questions and random effects for employees, which is more efficient than fixed

⁵ The plan was for each employee to receive five CEO answers. However, due to the implementation of the randomization procedure in Qualtrics, there was a higher likelihood of participants getting the CEO’s answer for questions 9 and 10. Consequently, 71.4% of employees ended up with exactly five CEO answers, while the rest received six. Overall, 52.9% of all answers provided to employees came from the CEO.

⁶ We did not adopt a multi-item scale for perceived helpfulness to ensure that a sufficient number of employees would participate in the study. Any additional item would imply 10 additional responses per participant.

effects, and true AI is orthogonal to employees. The effect of true AI is positive ($b = 0.824$; $p < .001$), which corresponds to an average increase of 17.6 percentage points in the probability that an answer is perceived as from the AI.⁷

[[INSERT TABLE 3 ABOUT HERE]]

Model 2 of Table 3 provides a linear regression with perceived helpfulness ($M = 3.621$, $SD = 1.005$) as the dependent variable. The model includes fixed effects for questions and for employees, because perceived AI is not orthogonal to employees. Standard errors are clustered by employee.⁸ The effect of perceived AI is negative ($b = -0.764$; $p < .001$), which gives a standardized mean difference of -0.760 . Thus, employees regarded answers they believed came from the CEO as more helpful than those they thought came from the AI. The effect of true AI is close to zero ($b = 0.020$; $p = .668$). Employees regarded answers that actually came from the CEO as equally helpful as those that actually came from the AI. We ran a mediation analysis with 10,000 bootstrap samples for the direct effect of true AI on perceived helpfulness and the indirect effect via perceived AI (Imai et al., 2010; Tingley et al., 2014). The direct effect is 0.021 (95% CI: $[-0.066, 0.109]$) and the indirect effect is -0.135 (95% CI: $[-0.191, -0.081]$). Perceived AI fully mediates the effect of true AI on perceived helpfulness.⁹

3.4. The Weak and Strong Form of Gen-AI Aversion

We posited that weak algorithm aversion relates to algorithmic performance and strong algorithm aversion is orthogonal to it. The CEO bot's task was to mimic the CEO. Hence, it performs better if employees have more difficulty detecting the communication's true source. Under weak Gen-AI aversion, employees' aversion to responses believed to come from AI would increase as CEO and AI responses are easier to tell apart (i.e., when detection is higher). Under strong Gen-AI aversion, employees' aversion to

⁷ This is $E[P(Y = 1 | \text{true AI} = 1)] - E[P(Y = 1 | \text{true AI} = 0)]$ and calculated as $1/J \sum_j E[P(Y = 1 | \text{true AI} = 1, \text{question} = j)] - 1/J \sum_j E[P(Y = 1 | \text{true AI} = 0, \text{question} = j)]$.

⁸ Clustering standard errors by employee and question yields similar results (see Table A2 in the Online Appendix).

⁹ The absolute magnitude of the indirect effect is small because of the first step of the multiplicative mediation chain (true AI \rightarrow perceived AI \rightarrow perceived helpfulness).

responses believed to come from AI would remain even if CEO and AI responses are difficult to tell apart (i.e., when detection is low). We investigate if the pattern of results fits with either.

Figure 3 shows detection and Gen-AI aversion by question. Detection is measured as accuracy or the proportion of times the source was correctly identified for a given question. For example, question 6 had the highest detection with an accuracy score of 0.705 and question 2 the lowest detection with an accuracy score of 0.495. Gen-AI aversion is measured as the difference in the perceived helpfulness between responses believed to come from AI and those from the CEO. Specifically, we regressed perceived helpfulness on perceived AI, interacted with each question (see Table A2 in the Online Appendix). The figure shows that as detection increases the difference in perceived helpfulness decreases, inconsistent with the weak form of Gen-AI aversion. Indeed, when detection is low, employees' aversion to responses believed to come from AI is high, consistent with the strong form of Gen-AI aversion.

[[INSERT FIGURE 3 ABOUT HERE]]

We further probe the relationship between detection and Gen-AI aversion through a regression. Model 1 of Table 4 provides a linear regression with perceived helpfulness as outcome variable. Fixed effects are included for question and employee, and standard errors are clustered by both. Question detection (i.e., accuracy by question) positively moderates the effect of perceived AI on perceived helpfulness ($b = 1.444, p = 0.044$). Thus, as in Figure 3, Gen-AI aversion is more pronounced when question detection is lower. The total effect of perceived AI is -0.905 (95% CI: $[-1.125, -0.685]$) for the lowest detection observed (question 2) and -0.602 (95% CI: $[-0.819, -0.385]$) for the highest detection observed (question 6). These findings indicate that Gen-AI aversion is greater for questions whose AI and CEO answers are harder to tell apart, which is consistent with the strong form of Gen-AI aversion.

[[INSERT TABLE 4 ABOUT HERE]]

3.4.1. Purpose of CEO communication

An additional way to explore the difference between the weak and strong form of Gen-AI aversion is to consider the purpose of CEO communication. Questions differ in their potential to elicit a response from the CEO revealing aspects of themselves (e.g., inner psychological processes or personality traits).

Because addressing such questions relies on personal information, it can be challenging to verify the validity of the response and consequently the response's source. Yet, in these situations, the source matters more for perceptions about the response's helpfulness. Therefore, under strong Gen-AI aversion, employees' aversion to responses believed to come from AI would be greater for questions with higher revelatory potential (even if CEO and AI responses are difficult to tell apart). In contrast, under weak Gen-AI aversion, employees' aversion to responses believed to come from AI would be less for potentially revelatory questions because CEO and AI responses are harder to tell apart. We investigate if the pattern of results fits with either.

We measure a question's revelatory potential using generative AI (for details, see section A.1. of the Online Appendix). We used eight LLMs: Claude 3 Haiku, Claude 3 Sonnet, Claude 3.5 Sonnet (from Anthropic), Llama 3.1 (from Meta), Mistral Large (from Mistral), GPT 3.5, GPT 4o-mini, and GPT 4o (from OpenAI). In the chain-of-thought prompt (Wei et al., 2022), we asked for a qualitative analysis of each question followed by a quantitative score on a scale of 0 (minimally revealing) to 10 (highly revealing). Each LLM assessed each question 10 times, with the scores averaged. The resulting scores are consistent across LLMs (Cronbach's alpha equals 0.971). The range of scores across questions is [2.89, 7.78].

Model 2 of Table 4 reports a linear regression with perceived helpfulness as the outcome variable. The specification is similar to model 1 but now includes an interaction between perceived AI and question revelatory potential. The interaction is negative ($b = -0.059$, $p = 0.053$), indicating that Gen-AI aversion is greater for questions with higher revelatory potential. Model 3 reports a logit regression of detection at the observation level, with random effects for employees. It includes question revelatory potential, without fixed effects for questions (because revelatory potential is constant by question). Question revelatory potential is associated with lower detection ($b = -0.119$, $p < 0.001$). Employees find it harder to correctly detect the source for potentially more revealing questions. These findings are consistent with the strong form of Gen-AI aversion because for questions with higher revelatory potential detection is low, yet aversion to responses believed to come from AI is high. In contrast, they are inconsistent with the

weak form of Gen-AI aversion because, under weak form, higher revelatory potential would link to higher detection and lower Gen-AI aversion.

3.5. Additional Analyses

As background, we provide additional analyses on the extent to which AI and CEO answers differed and on the extent to which accuracy and Gen-AI aversion differed across employees.

3.5.1. Differences between answers

The AI answers ($M = 134.6$, $SD = 36.2$) had more words than the CEO answers ($M = 83.3$, $SD = 38.2$; paired t -test: $t(9) = 2.927$, $p = .017$).

We compared the answers' content along five common CEO communication styles—excitable, stern, dramatic, rambling, and melancholy (Choudhury et al., 2019)—using Gen-AI (for details, see the section A.2 of the Online Appendix). We used eight LLMs: Claude 2.1, Claude 3 Haiku, Claude 3 Sonnet (from Anthropic), Llama 2 (from Meta), Mixtral 8x7B (from Mistral), GPT 3.5, GPT 4, and GPT 4 Turbo (from OpenAI). In the chain-of-thought prompt (Wei et al., 2022), we asked for a qualitative analysis of each answer followed by a quantitative score for each style on a scale of 0 (no presence) to 10 (maximum presence). Each LLM assessed each answer 10 times, with the scores averaged. The resulting scores are consistent across LLMs with Cronbach's alphas of .947 (excitable), .953 (stern), .952 (dramatic), .778 (rambling), and .963 (melancholy).

Figure 4 shows the scores averaged across LLMs for AI and CEO answers. The AI answers were rated as more excitable (4.23 vs. 2.61; paired t -test: $t(79) = 7.099$, $p < .001$) and dramatic (2.52 vs. 1.54; $t(79) = 5.032$, $p < .001$). The CEO answers were rated as more stern (2.43 vs. 3.33; $t(79) = -3.943$, $p < .001$). Only small differences existed for rambling (1.14 vs. 1.00; $t(79) = 2.057$, $p = .043$) and melancholy (0.82 vs. 0.89; $t(79) = -.864$, $p = .390$), with these styles rated low for both CEO and AI.

[[INSERT FIGURE 4 ABOUT HERE]]

3.5.2. Differences between employees

We analyzed the accuracy of perceived origin by subgroup of employees (see section A.3 of the Online Appendix). In terms of position, senior managers (.646) score higher than others (.583) (95% CI for difference: [-.020, .149]). In terms of tenure, accuracy for those who had three years or more (.619) was higher than for those with less than three years (.582) (95% CI for difference: [-.017, .123]). Thus, those more exposed to the CEO appear to be more able to differentiate between CEO and AI answers, yet even the most exposed do not exceed accuracy of 65% on average.

We also analyzed if the effect of perceived origin on perceived helpfulness differed by subgroup of employees. In a regression with interaction effects between perceived AI with seniority (managers) and another with tenure (three years or more), we find that the effect of perceived AI does not differ for senior managers ($b = .008, p = .970$). It is stronger for employees with longer tenure ($b = -.239, p = .107$), but not statistically so.

4. STUDY 2

Study 1 suggests the strong form of Gen-AI aversion exists. It revealed that employees regarded messages they believed were generated by AI as less helpful than those they thought came from the CEO, even when the two were *nearly indistinguishable*. Study 2 provides a causal test of the existence of the strong form of Gen-AI aversion. In this pre-registered study,¹⁰ we investigate whether people judge AI messages as less helpful than CEO messages, even if the messages are *identical*. Specifically, we manipulate a message's labeled origin while keeping its content identical. Furthermore, we use an audience less familiar with the CEO and therefore less likely to recognize AI and CEO messages as such.

4.1. Methods

4.1.1. Participants

US-based participants were recruited via Prolific, an online platform. Based on power calculations, the target sample was 210, to which a 20% margin was added for potential exclusions based on predefined

¹⁰ The pre-registration is available here: https://aspredicted.org/D2Z_52R

criteria.¹¹ We opened 252 positions and paid a fixed fee of \$2.05. We excluded 24 participants whose completion took less than two minutes or more than 15 minutes, and 12 participants who failed an attention check, yielding a sample size of 216 (57% female; M age = 37.3; SD age = 12.1). In the final sample, respondents' usage of AI tools (e.g., text-generation, image-generation, coding assistants) for personal use or work over the last seven days was 0 days (28.2%), 1-3 days (34.3%), 4-6 days (19.4%), or 7 days (19.1%). Employment status was full-time (55.9%), part-time (13.5%), or not (30.6%). In addition, 54.2% had ever been employed as a knowledge worker (someone whose job focuses on information, ideas, or data, rather than manual labor). Finally, 45.4% of respondents had traded a stock or fund in the past year, 25.0% had done so more than a year ago, and 29.6% had never traded.

4.1.2. Procedure

Participants had to assess the answers to four questions asked by financial analysts during Q1 2024 earnings calls. The questions were asked of the CEOs of Hershey, General Motors, Nvidia, and Amazon, and they were presented in the same order. To enhance generalizability (Shadish et al., 2002), we selected these companies because they differed in terms of industry and CEO characteristics (i.e., gender).¹² Each participant saw one randomly chosen answer per question in a 2 * 2 within subject design: actual origin (CEO / AI) and labeled origin (CEO / AI). For actual origin, the answer was from either the CEO or an LLM instructed to mimic the CEO in both content and style. The LLM was based on GPT 4o from OpenAI and given access to company documents (e.g., annual reports) and CEO interview transcripts (e.g., earnings calls) from 2023. The AI answers were obtained prior to inviting participants and retained unchanged, except for cropping where necessary to approximately match the length of the CEO's answer. For labeled origin, the answer was labeled as from either the CEO or an AI (Yin et al., 2024). The participants were told that the AI was trained on company information. The questions and answers are provided in the Online Appendix. The number of observations is $864 = 216$ (participants) * 4 (questions).

¹¹ Calculations are based on $\alpha = 0.05$, $\beta = 0.95$, and a standardized mean difference of 0.5, a medium effect size per Cohen (1992).

¹² Even though these companies were not selected for their similarity with Pulse, we can compare them using the same criteria of industry and CEO gender. Pulse's industry ("technology") is closest to Amazon and Pulse's CEO is male (as are the CEOs of Nvidia and Amazon).

4.1.3. Measures

We used two outcome variables. In line with study 1, perceived helpfulness was measured with the item “How helpful is this answer for making a decision on investing in the company?” (1 = not helpful at all, 5 = extremely helpful). In addition, share purchase intention was measured with the item “Based on this answer, how likely are you to buy the company’s shares in the next 12 months?” (1 = not likely at all, 5 = extremely likely) (Morwitz, 2014). A difference in perceived helpfulness or share purchase intention following responses labeled as AI compared to those labeled as CEO is an indicator of Gen-AI aversion.

4.2. Results

Model 1 of Table 5 reports a linear regression with perceived helpfulness ($M = 3.226$, $SD = 1.157$) as the dependent variable, with fixed effects for questions and random effects for participants. As expected, the effect of labeled AI (i.e., an indicator that is 1 when the answer was labeled as AI and 0 otherwise) is negative ($b = -0.146$; $p = .024$), yielding a standardized mean difference of -0.126 . Thus, answers labeled as from an AI were seen as less helpful than those labeled as from the CEO. In contrast, the effect of true AI (i.e., an indicator that is 1 when the answer is actually from the AI and 0 otherwise) is positive ($b = 0.879$; $p < .001$). Answers actually from the AI were perceived as more helpful. The findings for share purchase intention ($M = 2.819$, $SD = 1.197$) are similar (see model 2). The AI-labeled answers reduce share purchase intention ($b = -0.153$; $p = .016$), whereas actual AI answers increase it ($b = 0.675$; $p < .001$). Hence, labeling answers as from an AI reduces their potency, even as actual AI answers compare favorably to those of the CEO. Thus, these findings provide evidence for the strong form of Gen-AI aversion.

[[INSERT TABLE 5 ABOUT HERE]]

4.2.1. Additional analyses

First, we checked the manipulation. Participants were asked to recall whether the answer for the final company was provided by the CEO or the AI, and they tended to answer in line with the provided CEO or AI label (68.1%, $\text{Chi}^2(1) = 32.5$, $p < .001$). Second, we assessed whether the effect of labeled AI differed for actual CEO and AI answers. The effect of labeled AI is less negative for actual AI answers (perceived

helpfulness: $b = 0.088$, $p = 0.495$; share purchase intention: $b = 0.078$, $p = 0.541$) (see Table B1 in the Online Appendix).

5. DISCUSSION

This study employs both a field and a lab experiment to investigate whether a CEO bot can impersonate a human CEO in performing asynchronous communication. It also examines how the bot-generated communication is perceived by recipients. In the field experiment, we collaborated with a firm to develop a Gen-AI-based chatbot that mimics the communication content and style of the actual CEO. Using a randomized control trial (RCT), we study whether employees exhibit a weak or strong form of Gen-AI aversion (i.e., contingent on Gen-AI performance or not) towards the bot-generated communication. The two main findings are as follows: the CEO bot performs well on the impersonation task (i.e., it nearly passes the Wade Test). However, employees display a strong form of Gen-AI aversion towards communication that they believe is generated by the CEO bot, especially if such communication reveals personal aspects of the CEO. The lab experiment provides further evidence that general audience members performing the role of an investor exhibit a strong form of Gen-AI aversion towards communication generated by the CEO bot.

This study contributes to two research streams that are relevant to strategy and organizational scholars. The first is the growing literature on Gen-AI adoption and its impact on firms (e.g., Brynjolfsson et al., 2025; Eloundou et al., 2024). Prior studies show that Gen-AI can improve the productivity of individuals (Dell'Acqua et al., 2026; Noy & Zhang, 2023; Otis et al., 2024; Zhou & Lee, 2024), and may lead to drastic changes in how people operate within firms (Berg et al., 2023; Felten et al., 2021). We extend this literature by shifting the focus from employees and middle managers to the CEO, arguably the most consequential organizational actor. We focus on organizational communication, allowing us to explore the understudied impact of Gen-AI on an important organizational outcome and a main task for CEOs. Drawing on long-standing research in strategy that focuses on how CEOs can better employ technology to stay ahead of the “avalanche” of digital communication (e.g., Porter & Nohria, 2018), we

document both the potential of the personalized bot (i.e., nearly passing the Wade test) and its limitations (i.e., recipients exhibiting a strong form of Gen-AI aversion).

The second research stream is the established literature on algorithm aversion (e.g., Berger et al., 2021; Dietvorst et al., 2015; Jussupow et al., 2024). Extending past work that has extensively explored aversion to predictive algorithms, we present a conceptual framework of Gen-AI aversion (categorizing both weak and strong forms) and a typology of Gen-AI failures (categorizing three failure types). Crucially, we articulate conditions under which researchers can detect weak and strong aversion by pointing out how the weak and strong forms of aversion relate (or do not) to Gen-AI failures. Relatedly, we propose the Wade test to account for personalized Gen-AI technology and present the Wade test as a theoretical variant of the well-known Turing test.

Our study is not without limitations, offering opportunities for future research. First, the field experiment was run within a single firm, as is frequently the case for organizational field experiments (e.g., Choudhury et al., 2023; Gong & Png, 2024; Lane et al., 2022). Although we expect the conceptual framework to apply to other firms and settings, the effectiveness of a CEO bot in other circumstances may vary depending on the quantity and quality of the available training data, the personality of the CEO, and firm-level characteristics such as firm size, geographic location(s), and geographic distribution of the firm. For example, the CEO bot in the first experiment was trained on a large set of textual data from the CEO's communication history both internal and external to the firm. Consequently, the CEO bot in our study did well in the Wade test. However, it is plausible that a similar personalized bot developed for another CEO may not have access to high-quality training data, possibly because the CEO was recently appointed or due to a lack of digitized communication. In that case, the resulting CEO bot may struggle to mimic the CEO.

Additionally, the CEO bot deployed in our study was developed in early 2024. It is conceivable that personalized bots will become even more powerful, given the rapid advancement of Gen-AI in new capabilities, including reasoning (e.g., Huang et al., 2022). Exciting recent developments include retrieval augmented generation (RAG) and online search capabilities, both of which help LLMs generate more

context-aware responses. RAG allows an LLM to retrieve information from a curated and frequently updated set of documents (e.g., market reports or competitive intelligence) at generation time, rather than relying solely on information available at training time. Similarly, LLMs with online search capabilities can access real-time information, making it possible to tailor messages to the latest developments inside and outside the firm. Even as LLMs can increasingly account for recent information, to what extent can they be used for forward-looking insights as required for strategic decision making? An active line of inquiry is whether LLMs have, or can produce, strategic foresight. Early evidence indicates that they may. For example, Doshi et al. (2025) document congruence between LLMs and human experts in predicting business model success. Furthermore, strategic foresight depends not only on forecasting outcomes but also on the ability to analyze and reason, which LLMs increasingly can (Zhang et al., 2024).

Second, both of our experiments focus on the immediate usage of Gen-AI. They, therefore, do not allow us to track how effectively CEO bots can facilitate CEO communication in the long run. Prior studies on technology management suggest that it is possible for the weak form of aversion to reduce after repeated exposure to a new technology (e.g., Filiz et al., 2021). However, given that we report a strong form of aversion towards the CEO bot, it remains to be studied whether and how such aversion might be mitigated through longer-term exposure to CEO bots.

Third, the field experiment's participation rate of about 13% of employees of the firm is within the range of existing field experiments (e.g., Choudhury et al., 2023; Gong & Png, 2024; Guzman et al., 2020) but nonetheless modest. This is because employees participated in the experiment voluntarily after seeing participation recruiting messages in the internal Slack channel. This raises the possibility of selection into the experiment based on the employee's availability, interest in Gen-AI models, and interest in communicating with the CEO. It is possible that employees who were more interested in Gen-AI, self-selected to participate in the experiment. However, it is also plausible that self-selection by employees who are interested in Gen-AI leads to more conservative estimates of Gen-AI aversion. If all employees, including those who are less familiar with Gen-AI, had participated in the experiment, the degree of the strong form aversion might have been even greater.

Our findings open several avenues for future research on Gen-AI and the performance of tasks by top managers, such as the CEO. Researchers could study whether communication practices such as adding an electronic signature or a disclaimer from the CEO to the bot-generated content change patterns of Gen-AI aversion. The strategy literature (e.g., Kaplan, 2008) has long documented the practice of adding the signature of the CEO to asynchronous communication prepared by other humans for company annual reports. In the case of the bot, it would be interesting to study whether appending the CEO signature to bot generated content or the CEO reviewing the bot generated content alters patterns of Gen-AI aversion. In a similar spirit, researchers could examine how audiences respond differently to bot-generated content compared to messages prepared by other humans supporting CEO communication. Both approaches involve content not directly authored by the CEO. However, the bot uniquely attempts to emulate the CEO's personal style, whereas human writers typically do not have this explicit directive. It would be interesting to understand the implications of these similarities and differences.

Researchers could go beyond LLMs to examine multimodal Gen-AI communication, including speech and video, building on recent examples of leaders using Gen-AI technologies to remotely “give” speeches (Zhuang, 2024). Relatedly, this study compared human or Gen-AI communication. Possibilities also exist to study how CEOs can augment their communication or decision making with Gen-AI tools. Researchers could also estimate using a large sample of firms whether and to what extent, employing the CEO bot frees up CEO time to perform other tasks, what alternate tasks CEOs performed with the time saved, and what organizational outcomes follow. Crucially, studies could focus on whether and to what extent CEO bots might replace human administrative assistants who help CEOs with asynchronous communication.

In addition to CEOs, future research could examine how personalized bots might effectively help other upper-echelon members, such as CFOs, with their asynchronous communication. Such studies could measure how communication purposes vary across C-suite roles. For instance, CFOs may prioritize conveying decisions rather than motivating employees or sharing personal stories. Researchers could then test whether the effectiveness of personalized bots differs across roles. Likewise, scholars could study

nonprofit or public-sector leaders, whose communication goals may differ from those of corporate executives. Beyond top managers such as the CEO, future studies could explore how personalized bots might affect communication by employees and middle managers. Recent reporting (e.g., David, 2024) indicates that firms like Otter.ai are developing personalized Gen-AI assistants that employees can delegate to perform communication and join meetings on their behalf. This raises the potential to study how personalized bots might affect organizational communication more broadly and what implications this might have for organizational and knowledge hierarchies (e.g., Garicano, 2000).

Our results have practical implications for managers. In an interview, Eric Yuan (Zoom founder) outlined the promise of personalized bots: *“one of the big benefits of AI at work will be letting us all create something he calls a “digital twin”—essentially a [copy] avatar of yourself that can go to Zoom meetings on your behalf and even make decisions for you while you spend your time on more important things, like your family.”* (Patel, 2024). However, despite this ambitious vision, there exists a background of cautionary notes on the effectiveness of personalized bots: *“There is also an argument that A.I. won’t truly be useful until it really understands its users — their communication styles, their likes and dislikes.”* (Roose, 2023). Our study sheds light on both the promise and the limitations of personalized bots for CEOs. Gen-AI tools can help CEOs generate communication that can closely mimic their own, yet CEOs should weigh the purpose of each communication before delegating it. Gen-AI can assist, but employees may still prefer to hear from the CEO for certain purposes, not a Gen-AI that mimics the CEO.

In conclusion, this study explores the use of Gen-AI for communication by CEOs. The Pulse CEO, an early experimenter with Gen-AI, has now fully integrated Gen-AI in his workflow. Several other CEOs have followed suit. Jensen Huang (Nvidia) has a personal AI tutor available to him all of the time. He describes a feeling of empowerment and the potential for productivity gains (Abram, 2025). Satya Nadella (Microsoft) uses a Gen-AI co-pilot during meetings with the CEOs of customers (Millward, 2024). Similarly, Lisa Su (AMD) encourages the use of AI copilots throughout the organization and personally experiments with them for her workflow (Ignatius, 2025). This momentum presents a wealth

of research opportunities for scholars to understand the role and impact of Gen-AI on CEOs and their activities.

REFERENCES

- Abram, C. (2025). *NVIDIA CEO Jensen Huang's vision for the future* [Video]. YouTube.
<https://www.youtube.com/watch?v=7ARBJQn6QkM>.
- Akstinaite, V., Robinson, G., & Sadler-Smith, E. (2020). Linguistic markers of CEO hubris. *Journal of Business Ethics*, *167*, 687–705.
- Allen, R., & Choudhury, P. (2022). Algorithm-augmented work and domain experience: The countervailing forces of ability and aversion. *Organization Science*, *33*(1), 149–169.
- Bandiera, O., Prat, A., Hansen, S., & Sadun, R. (2020). CEO behavior and firm performance. *Journal of Political Economy*, *128*(4), 1325–1369.
- Berg, J. M., Raj, M., & Seamans, R. (2023). Capturing value from artificial intelligence. *Academy of Management Discoveries*, *9*(4), 424–428.
- Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). Watch me improve—Algorithm aversion and demonstrating the ability to learn. *Business & Information Systems Engineering*, *63*(1), 55–68.
- Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*, 21–34.
- Bluvstein, S., Zhao, X., Barasch, A., & Schroeder, J. (2024). Imperfectly human: The humanizing potential of (corrected) errors in text-based communication. *Journal of the Association for Consumer Research*, *9*(3), 332–343.
- Bojić, L., Zagovora, O., Zelenkauskaitė, A., Vuković, V., Čabarkapa, M., Veseljević Jerković, S., & Jovančević, A. (2025). Comparing large language models and human annotators in latent content analysis of sentiment, political leaning, emotional intensity and sarcasm. *Scientific Reports*, *15*(1), 11477.
- Boyle, M. (2025). Meetings won't be the same when the boss sends an AI bot. *Bloomberg*.
<https://www.bloomberg.com/news/features/2025-04-15/meetings-won-t-be-the-same-when-the-ceo-sends-an-ai-bot>.
- Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics*, *140*(2), 889–942.
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv preprint arXiv:2303.12712*.
- Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239.
- Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825.

- Cha, S. E., Hewlin, P. F., Roberts, L. M., Buckman, B. R., Leroy, H., Steckler, E. L., Ostermeier, K., & Cooper, D. (2019). Being your true self at work: Integrating the fragmented research on authenticity in organizations. *Academy of Management Annals*, *13*(2), 633–671.
- Chen, Z., & Chan, J. (2024). Large language model in creative work: The role of collaboration modality and user expertise. *Management Science*, *70*(12), 9101–9117.
- Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. de O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., Ryder, N., Pavlov, M., Power, A., Kaiser, L., Bavarian, M., Winter, C., Tillet, P., Petroski Such, F., Cummings, D., Plappert, M., ... Zaremba, W. (2021). Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*.
- Choudhury, P., Khanna, T., & Sevchenko, V. (2023). Firm-induced migration paths and strategic human-capital outcomes. *Management Science*, *69*(1), 419–445.
- Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. (2019). Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal*, *40*(11), 1705–1732.
- Clapp-Smith, R., Vogelgesang, G. R., & Avey, J. B. (2009). Authentic leadership and positive psychological capital: The mediating role of trust at the group level of analysis. *Journal of Leadership & Organizational Studies*, *15*(3), 227–240.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.
- Dargnies, M.-P., Hakimov, R., & Kübler, D. (2026). Aversion to hiring algorithms: Transparency, gender profiling, and self-confidence. *Management Science*, *72*(1), 285–301.
- David, E. (2024). Otter is making AI bots part of the group chat. *The Verge*.
<https://www.theverge.com/2024/2/13/24071269/otter-generative-ai-meeting-transcripts-group-chat>.
- Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Kraymer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Organization Science*. Forthcoming.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170.
- Doshi, A. R., Bell, J. J., Mirzayev, E., & Vanneste, B. S. (2025). Generative artificial intelligence and evaluating strategic decisions. *Strategic Management Journal*, *46*(3), 583–610.
- Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), eadn5290.
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, *384*(6702), 1306–1308.

- Eklund, J. C., & Mannor, M. J. (2026). Curious and analytical: How analysts evaluate and respond to executive communications about firm strategy. *Strategic Management Journal*. Forthcoming.
- Felten, E., Raj, M., & Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 42(12), 2195–2217.
- Filiz, I., Judek, J. R., Lorenz, M., & Spiwoks, M. (2021). Reducing algorithm aversion through experience. *Journal of Behavioral and Experimental Finance*, 31, 100524.
- Frisch, I., & Giulianelli, M. (2024). LLM agents in interaction: Measuring personality consistency and linguistic alignment in interacting populations of large language models. *arXiv preprint arXiv:2402.02896*.
- Gaessler, F., & Piezunka, H. (2023). Training with AI: Evidence from chess computers. *Strategic Management Journal*, 44(11), 2724–2750.
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*.
- Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, 108(5), 874–904.
- George, B., Sims, P., McLean, A. N., & Mayer, D. (2007). Discovering your authentic leadership. *Harvard Business Review*, 85(2), 129–138.
- Gong, J., & Png, I. P. L. (2024). Automation enables specialization: Field evidence. *Management Science*, 70(3), 1580–1595.
- Guzman, J., Oh, J. J., & Sen, A. (2020). What motivates innovative entrepreneurs? Evidence from a global field experiment. *Management Science*, 66(10), 4808–4819.
- Hambrick, D. C. (2007). Upper echelons theory: An update. *Academy of Management Review*, 32(2), 334–343.
- Hambrick, D. C., & Quigley, T. J. (2014). Toward more accurate contextualization of the CEO effect on firm performance. *Strategic Management Journal*, 35(4), 473–491.
- Helfat, C. E., & Peteraf, M. A. (2015). Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic Management Journal*, 36(6), 831–850.
- Huang, J., & Chang, K. C.-C. (2022). Towards reasoning in large language models: A survey. *arXiv preprint arXiv:2212.10403*.
- Ignatius, A. (2025). AMD's Lisa Su on experimenting with AI. *HBR On Strategy* (Episode 111) [Audio podcast]. Harvard Business Review.
<https://hbr.org/podcast/2025/05/amds-lisa-su-on-experimenting-with-ai>.
- Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15(4), 309–334.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38.

- Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. *Academy of Management Journal*, 67(1), 5–32.
- Jiang, L., John, L. K., Boghrati, R., & Kouchaki, M. (2022). Fostering perceptions of authenticity via sensitive self-disclosure. *Journal of Experimental Psychology: Applied*, 28(4), 898–915.
- Jussupow, E., Benbasat, I., & Heinzl, A. (2024). An integrative perspective on algorithm aversion and appreciation in decision-making. *MIS Quarterly*, 48(4), 1575–1590.
- Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R. (2023). Challenges and applications of large language models. *arXiv preprint arXiv:2307.10169*.
- Kaplan, S. (2008). Cognition, capabilities, and incentives: Assessing firm response to the fiber-optic revolution. *Academy of Management Journal*, 51(4), 672–695.
- Kim, Y., Lee, J., Kim, S., Park, J., & Kim, J. (2024). Understanding users' dissatisfaction with ChatGPT responses: Types, resolving tactics, and the effect of knowledge level. In *Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24)* (pp. 385–404).
- Kotter, J. P. (1990). What leaders really do. *Harvard Business Review*, 68(3), 103–111.
- Lane, J. N., Teplitzkiy, M., Gray, G., Ranu, H., Menietti, M., Guinan, E. C., & Lakhani, K. R. (2022). Conservatism gets funded? A field experiment on the role of negative information in novel project evaluation. *Management Science*, 68(6), 4478–4495.
- Lehman, D. W., O'Connor, K., Kovács, B., & Newman, G. E. (2019). Authenticity. *Academy of Management Annals*, 13(1), 1–42.
- Leong, J., Tang, J., Cutrell, E., Junuzovic, S., Baribault, G. P., & Inkpen, K. (2024). Dittos: Personalized, embodied agents that participate in meetings when you are unavailable. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2), 1–28.
- Liu, E. H., Chambers, C. R., & Moore, C. (2023). Fifty years of research on leader communication: What we know and where we are going. *Leadership Quarterly*, 34(6), 101734.
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103.
- Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4), 629–650.
- Microsoft. (2024). *2024 Annual Work Trend Index*.
<https://www.microsoft.com/en-us/worklab/work-trend-index/ai-at-work-is-here-now-comes-the-hard-part>.
- Millward, W. T. (2024). Microsoft's Nadella at Ignite 2024: "We are a platform-first and a partner-first company." *CRN*.
<https://www.crn.com/news/ai/2024/microsoft-ignite-2024-ceo-nadella-calls-ai-a-tremendous-partner-opportunity>.
- Morwitz, V. (2014). Consumers' purchase intentions and their behavior. *Foundations and Trends in Marketing*, 7(3), 181–230.
- Murphy, K. P. (2023). *Probabilistic machine learning: Advanced topics*. MIT Press.

- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192.
- Önköl, D., Goodwin, P., Thomson, M., Gönöl, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, 22(4), 390–409.
- Otis, N., Clarke, R., Delecourt, S., Holtz, D., & Koning, R. (2024). The uneven impact of generative AI on entrepreneurial performance. SSRN Working Paper 4671369.
- Patel, N. (2024). Zoom CEO Eric Yuan wants AI clones in meetings. *The Verge*.
<https://www.theverge.com/2024/6/3/24168733/zoom-ceo-ai-clones-digital-twins-videoconferencing-decoder-interview>.
- Porter, M. E., & Nohria, N. (2018). How CEOs manage time. *Harvard Business Review*, 96(4), 42–51.
- Rego, A., Reis Júnior, D., & Cunha, M. P. (2015). Authentic leaders promoting store performance: The mediating roles of virtuousness and potency. *Journal of Business Ethics*, 128(3), 617–634.
- Roose, K. (2023). Personalized A.I. agents are here: Is the world ready for them? *The New York Times*.
<https://www.nytimes.com/2023/11/10/technology/personalized-ai-agents.html>.
- Ross, S. I., Martinez, F., Houde, S., Muller, M., & Weisz, J. D. (2023). The programmer's assistant: Conversational interaction with a large language model for software development. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)* (pp. 491–514).
- Schick, T., & Schütze, H. (2021). It's not just size that matters: Small language models are also few-shot learners. *arXiv preprint arXiv:2009.07118*.
- SimanTov-Nachlieli, I. (2025). More to lose: The adverse effect of high performance ranking on employees' preimplementation attitudes toward the integration of powerful AI aids. *Organization Science*, 36(1), 1–20.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Cengage Learning.
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, 59(5), 1–38.
- Tong, S., Jia, N., Luo, X., & Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, 42(9), 1600–1631.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460.
- Vanneste, B. S., & Puranam, P. (2025). Artificial intelligence, trust, and perceptions of agency. *Academy of Management Review*, 50(4), 726–744.
- Wang, D. S., & Hsieh, C. C. (2013). The effect of authentic leadership on employee trust and employee engagement. *Social Behavior and Personality*, 41(4), 613–624.
- Wang, H., Sui, Y., Luthans, F., Wang, D., & Wu, Y. (2014). Impact of authentic leadership on performance: Role of followers' positive psychological capital and relational processes. *Journal of Organizational Behavior*, 35(1), 5–21.

- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824–24837.
- Yates, J., & Orlikowski, W. J. (1992). Genres of organizational communication: A structural approach to studying communication and media. *Academy of Management Review*, 17(2), 299–326.
- Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, 32(4), 403–414.
- Yin, Y., Jia, N., & Wakslak, C. J. (2024). AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences*, 121(14), e2319112121.
- Zhang, T., Qiu, L., Guo, Q., Deng, C., Zhang, Y., Zhang, Z., Zhou, C., Wang, X., & Fu, L. (2023). Enhancing uncertainty-based hallucination detection with stronger focus. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)* (pp. 915–932).
- Zhang, Y., Mao, S., Ge, T., Wang, X., de Wynter, A., Xia, Y., Wu, W., Song, T., Lan, M., & Wei, F. (2024). LLM as a mastermind: A survey of strategic reasoning with large language models. *arXiv preprint arXiv:2404.01230*.
- Zhang, Y., Tuk, M. A., & Klesse, A. K. (2024). Giving AI a human touch: Highlighting human input increases the perceived helpfulness of advice from AI coaches. *Journal of the Association for Consumer Research*, 9(3), 344–356.
- Zhou, E., & Lee, D. (2024). Generative artificial intelligence, human creativity, and art. *PNAS Nexus*, 3(3), pgae052.
- Zhuang, Y. (2024). Imran Khan's 'victory speech' from jail shows A.I.'s peril and promise. *The New York Times*.
<https://www.nytimes.com/2024/02/11/world/asia/imran-khan-artificial-intelligence-pakistan.html>.

TABLE 1. Questions for the CEO and AI (Study 1)

Number	Theme	Question
1	Strategy and future planning	Considering the rapid advancements in AI and competitive pressures, how does Pulse plan to adapt its strategy to maintain its growth and market position over the next five years?
2	Company culture and values	How do you ensure that Pulse’s core values are lived out in the day-to-day operations and decision-making processes at the company, especially as it continues to grow?
3	Founders’ experiences and decisions	Looking back on the early days of Pulse, what is one decision you believe was critical to the company’s success, and what lesson did it teach you about entrepreneurship?
4	Product development and innovation	As [...], how is Pulse positioning its products to lead in this space while also ensuring they meet the evolving needs of users?
5	Challenges and mistakes	Can you share a significant challenge Pulse faced in its growth journey and how you navigated it, including any mistakes made and the learnings taken from that experience?
6	Market position and competitive landscape	Who do you see as Pulse’s main competitors in the [...] industry, and how do you plan to differentiate and maintain a competitive advantage?
7	Customer impact and use cases	Can you share an example of a particularly innovative or impactful way that a customer has utilized Pulse to transform their business or daily operations?
8	Team and personal development	How does Pulse foster a culture of continuous learning and development among its team members, and can you share a personal growth story from your journey within the company?
9	Global expansion and localization	[...]
10	[...]	10. What have been some of the most challenging aspects of [...] at scale, and how has Pulse overcome these challenges to ensure a seamless user experience?

Note. Original text in brackets omitted to preserve anonymity.

TABLE 2. Confusion matrix of answer’s origin (Study 1)

		Actual origin	
		AI	CEO
Perceived origin	AI	26.7%	20.6%
	CEO	20.5%	32.3%

Note. $N = 1,050$.

TABLE 3. Regressions of perceived AI and perceived helpfulness (Study 1)

	(1) Perceived AI		(2) Perceived helpfulness	
	<i>b</i>	<i>se</i>	<i>b</i>	<i>se</i>
Perceived AI			-0.764	0.063
True AI	0.824	0.136	0.020	0.047
Questions	FE		FE	
Employees	RE		FE	
Constant	0.322	0.218	3.957	0.079
<i>n</i> observations	1,050		1,050	
<i>n</i> employees	105		105	
<i>Chi</i> ²	134.0			
<i>LL</i>	-640.8			
<i>F</i>			35.2	
<i>R</i> ²			0.182	

Note. RE is random effects and FE is fixed effects. Model 1 is a logistic regression. Model 2 is a linear regression with standard errors clustered by employee.

TABLE 4. Regressions assessing the strong form of Gen-AI aversion (Study 1)

	(1) Perceived helpfulness		(2) Perceived helpfulness		(3) Detection	
	<i>b</i>	<i>se</i>	<i>b</i>	<i>se</i>	<i>b</i>	<i>se</i>
Perceived AI	-1.619	0.379	-0.471	0.148		
True AI	0.006	0.060	0.011	0.059	-0.163	0.129
Perceived AI * question detection	1.444	0.618				
Perceived AI * question revelatory potential			-0.059	0.027+		
Question revelatory potential					-0.119	0.035
Questions	FE		FE		No	
Employees	FE		FE		RE	
Constant	3.982	0.043	3.970	0.042	1.054	0.198
<i>n</i> observations	1,050		1,050		1,050	
<i>n</i> employees	105		105		105	
<i>F</i>	39.0		39.5			
<i>R</i> ²	0.614		0.614			
<i>Chi</i> ²					14.0	
<i>LL</i>					-701.9	

Note. FE denotes fixed effects and RE is random effects. Models 1 and 2 are linear regressions with standard errors clustered by employee and question. Model 3 is a logistic regression. Question detection (model 1) and question revelatory potential (model 2) are constant by question, and therefore are accounted for via question fixed effects. Question detection is the accuracy or the proportion of times the source was correctly identified for a given question. Question revelatory potential is the degree to which a

question seeks to elicit a response that reveals aspects of the CEO themselves. The outcome variable in model 3, detection, is at the level of the observation (i.e., an employee correctly identifying or not the source of a communication).

TABLE 5. Regressions of perceived helpfulness and purchase intention (Study 2)

	(1)		(2)	
	Perceived helpfulness		Share purchase intention	
	<i>b</i>	<i>se</i>	<i>b</i>	<i>se</i>
Labeled AI	-0.146	0.065	-0.153	0.064
True AI	0.879	0.065	0.675	0.064
Companies	FE		FE	
Participants	RE		RE	
Constant	3.080	0.085	2.855	0.089
<i>n</i> observations	864		864	
<i>n</i> participants	216		216	
<i>Chi</i> ²	244.1		175.8	
<i>R</i> ²	0.190		0.124	

Note. FE is fixed effects and RE is random effects. Models 1 and 2 are linear regressions.

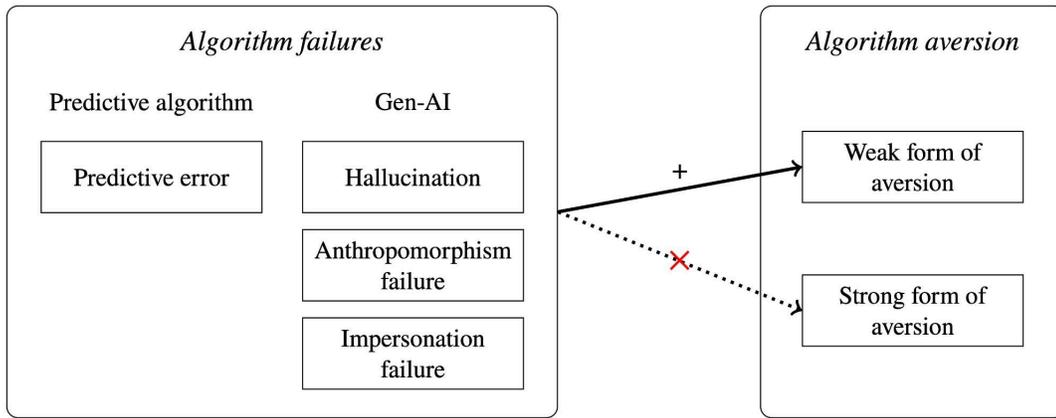


FIGURE 1. Relationship between algorithm failures and forms of algorithm aversion

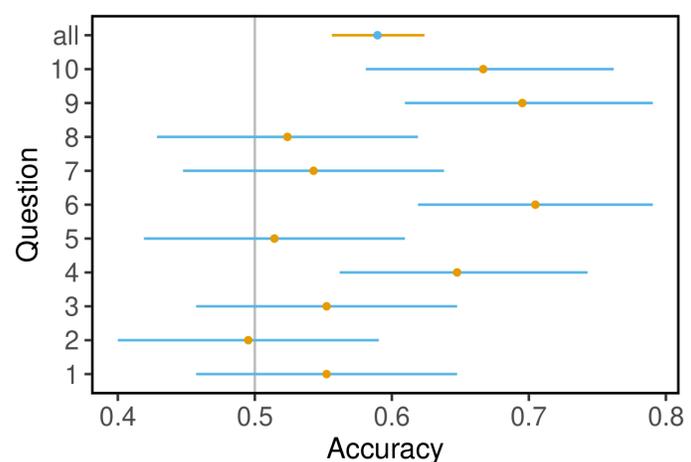


FIGURE 2. Accuracy of perceived origin (Study 1)

Note. Bootstrapped 95% CIs are shown, each based on 10,000 samples. For all questions combined (“all”), bootstrapping is at the employee level to take into account the nested data structure.

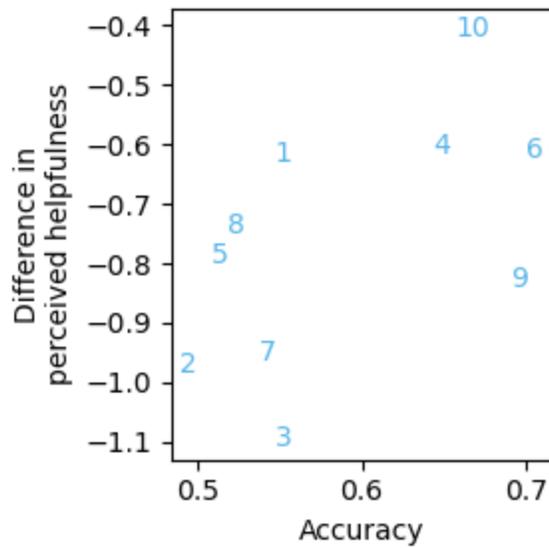


FIGURE 3. Accuracy and difference in perceived helpfulness per question (Study 1)

Note. The correlation is .555. Difference in perceived helpfulness captures a negative view of communication believed to come from AI compared to those from the CEO. Gen-AI aversion is greater as this difference is more negative. Greater accuracy indicates that employees are better able to detect the communication’s true origin.

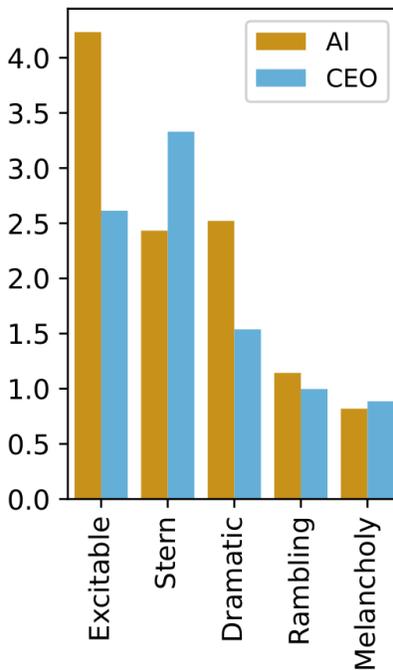


FIGURE 4. Average scores of CEO and AI answers on five communication styles (Study 1)

Note. Each bar represents the average of 10 answers assessed by 8 LLMs.

ONLINE APPENDIX

A. STUDY 1

A.1. Questions

A.2. Answers

A.3. Employees

B. STUDY 2

B1. CEO and AI answers

B2. Additional analyses

REFERENCES

A. STUDY 1

A.1. Questions

Table A1 provides the values used for Figure 2.

TABLE A1. Accuracy of perceived origin (Study 1)

Question	Mean	95% CI low	95% CI high
1	0.552	0.457	0.648
2	0.495	0.400	0.590
3	0.552	0.457	0.648
4	0.648	0.562	0.743
5	0.514	0.419	0.610
6	0.705	0.619	0.790
7	0.543	0.448	0.638
8	0.524	0.429	0.619
9	0.695	0.610	0.790
10	0.667	0.581	0.762
all	0.590	0.556	0.624

Note. Bootstrapped 95% CIs are shown, each based on 10,000 samples. For all questions combined (“all”), bootstrapping is at the employee level to take into account the nested data structure.

Table A2 (model 2) provides the values used for Figure 3.

TABLE A2. Difference in perceived helpfulness per question (Study 1)

	(1)		(2)	
	Perceived helpfulness		Perceived helpfulness	
	<i>b</i>	<i>se</i>	<i>b</i>	<i>se</i>
Perceived AI	-0.764	0.081		
True AI	0.020	0.060	0.007	0.061
Perceived AI * Q1			-0.615	0.105
Perceived AI * Q2			-0.971	0.092
Perceived AI * Q3			-1.095	0.067
Perceived AI * Q4			-0.603	0.084
Perceived AI * Q5			-0.786	0.125
Perceived AI * Q6			-0.609	0.070
Perceived AI * Q7			-0.950	0.088
Perceived AI * Q8			-0.736	0.093
Perceived AI * Q9			-0.825	0.080
Perceived AI * Q10			-0.406	0.084
Questions	FE		FE	
Employees	FE		FE	
Constant	3.972	0.046	3.987	0.032
<i>n</i> observations	1,050		1,050	
<i>n</i> employees	105		105	
<i>F</i>	44.8			

R^2 0.612 0.620

Note. FE is fixed effects. Models 1 and 2 are linear regressions. Standard errors are clustered by employee and question.

A.1.1. Large language models (LLMs)

We used generative artificial intelligence (AI) to assess a question’s revelatory potential (Liu et al., 2023). We employed the eight LLMs shown in Table A3.

TABLE A3. Summary of LLMs used for evaluating question revelatory potential (Study 1)

Name	Developer	Version	Release date
Claude 3 Haiku	Anthropic	claude-3-haiku-20240307	7 March 2024
Claude 3 Sonnet	Anthropic	claude-3-sonnet-20240229	29 February 2024
Claude 3.5 Sonnet	Anthropic	claude-3-5-sonnet-20241022	22 October 2024
Llama 3.1	Meta	llama3-1-70b-instruct-v1	23 July 2024
Mistral Large	Mistral	mistral-large-2402-v1	26 February 2024
GPT 3.5	OpenAI	gpt-3.5-turbo-0125	25 January 2024
GPT 4o-mini	OpenAI	gpt-4o-mini-2024-07-18	18 July 2024
GPT 4o	OpenAI	gpt-4o-2024-08-06	6 August 2024

A.1.2. Prompts

We followed a chain-of-thought prompt (Wei et al., 2022), which started with a system message (a general instruction to the AI) and then alternated between human and AI messages. The dynamic text <question> refers to the question posed to the CEO and AI, with the company’s name changed to simply “company”. The requested JSON format is a standard format for representing structured data.

System: As a social scientist specializing in textual analysis, your task is to analyze a question related to an anonymous company and evaluate the degree to which it seeks to elicit a response from the CEO that reveals aspects of themselves (e.g., inner psychological processes or personality traits).

Human: Carefully read the question provided below. Your task is to qualitatively assess the extent to which this question seeks to elicit a response from the CEO that reveals less obvious or hidden aspects of themselves. Reflect on the degree to which the question requires the CEO to disclose their personality, motivations, or inner thought processes. Indicate whether the question inherently promotes a revelatory and introspective response or an impersonal one. You don’t need to provide your assessment in a numerical form yet; just form your opinion on the level of revelatory potential of the question to the company.

The question posed to the CEO is:

<question>

AI: <response>

Human: Based on your qualitative assessment of the question’s ability to reveal less obvious or hidden aspects of the CEO, now quantitatively score its revelatory potential on a scale from 0 to 10. A score of 0 indicates that the question is minimally revealing and a response is unlikely to provide meaningful or unique insights into the CEO. A score of 10 signifies that the question is highly revealing and a response

is likely to provide meaningful or unique insights into the CEO. Ensure that your score is represented accurately in a JSON object, with “question” as a key and the assigned score as its value.

AI: <response>

A.1.3. Settings

Each LLM assessed each question ten times, with the scores averaged. Before averaging, each LLM provided 10 (questions) * 10 (repeats) = 100 assessments. After averaging, each LLM yielded one score for each question.

Temperature is a setting that controls the randomness of the response. It was 0.5 for the first AI response (to allow the qualitative analysis to highlight different aspects) and 0 for the second AI response (to ensure the quantitative analysis returned its most confident response and in the correct JSON format).

A.1.4. Results

TABLE A4. Cronbach’s alpha for question revelatory potential by LLM (Study 1)

LLM	Cronbach’s alpha
Claude 3 Haiku	0.932
Claude 3 Sonnet	0.989
Claude 3.5 Sonnet	0.999
Llama 3.1	0.992
Mistral	0.985
GPT3.5	0.948
GPT4o mini	0.975
GPT4o	0.996

TABLE A5. Cronbach’s alpha for question revelatory potential across LLMs (Study 1)

LLMs	Cronbach’s alpha
8	0.971

Note. The Cronbach’s alphas shown here are for the measure used (i.e., after averaging assessments by LLM).

TABLE A6. Scores of question revelatory potential (Study 1)

Question	mean	s.d.
1	3.21	1.42
2	6.29	1.50
3	7.35	0.74
4	3.35	1.69
5	7.78	0.65
6	4.00	1.81
7	3.99	1.81
8	7.36	0.91

9	2.89	1.05
10	3.94	1.44

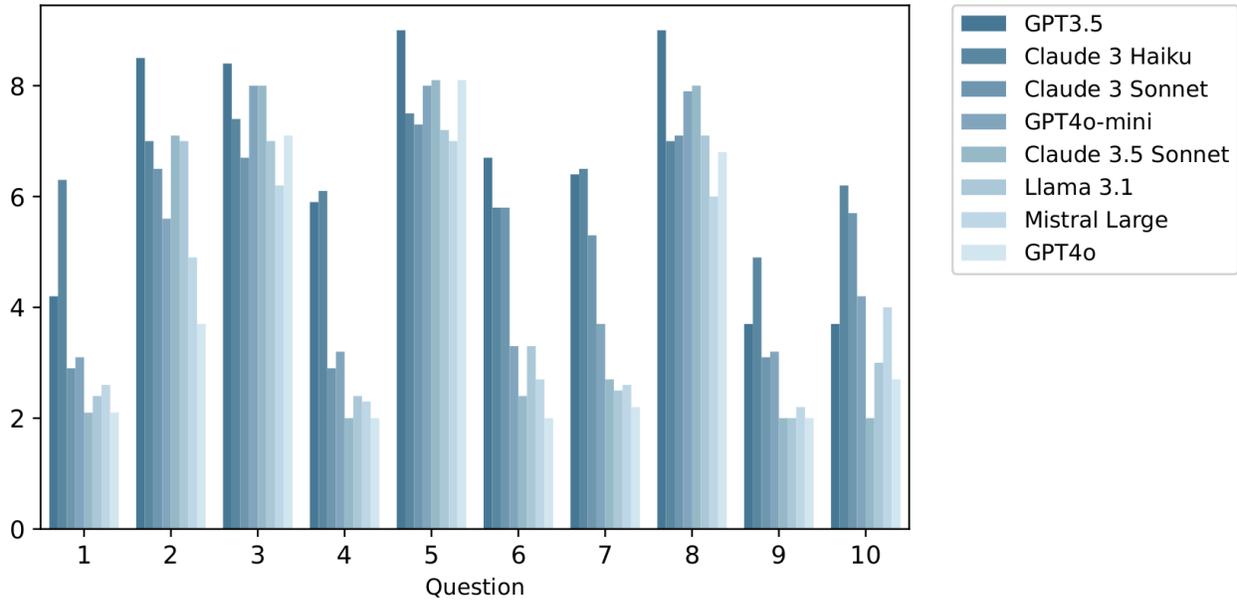


FIGURE A1. Average scores of question revelatory potential by LLM (Study 1)

Note. Each bar represents the average of 10 answers assessed by 1 LLM.

A.2. Answers

A.2.1. Large language models (LLMs)

We used generative artificial intelligence (AI) to assess five CEO communication styles—excitable, stern, dramatic, rambling, and melancholy (Choudhury et al., 2019)—of the CEO and AI answers. We employed the eight LLMs shown in Table A7.

TABLE A7. Summary of LLMs used for evaluating communication styles (Study 1)

Name	Developer	Version	Release date
Claude 2.1	Anthropic	claude-2.1	21 November 2023
Claude 3 Haiku	Anthropic	claude-3-haiku-20240307	7 March 2024
Claude 3 Sonnet	Anthropic	claude-3-sonnet-20240229	29 February 2024
Llama 2	Meta	llama2-70b-chat	18 July 2023
Mixtral 8x7B	Mistral	mixtral-8x7b-instruct-v0.1	11 December 2023
GPT 3.5	OpenAI	gpt-3.5-turbo-0125	25 January 2024
GPT 4	OpenAI	gpt-4-0613	13 June 2023
GPT 4 Turbo	OpenAI	gpt-4-0125-preview	25 January 2024

A.2.2. Prompts

We followed a chain-of-thought prompt (Wei et al., 2022), which started with a system message (a general instruction to the AI) and then alternated between human and AI messages. The dynamic text <answer>

refers to the answers provided by the CEO and AI. The requested JSON format is a standard format for representing structured data.

System: As a social scientist specializing in textual analysis, your task is to conduct a qualitative assessment of a text’s style, followed by quantifying this assessment into specific categories.

Human: Carefully read the text provided below. Your task is to qualitatively assess the style of the writing. Consider how excitable, stern, dramatic, rambling, and melancholy the text is. Think about the word choice, sentence structure, and overall tone. You don’t need to provide your assessment in a numerical form yet; just form your opinion on the prominence of each style within the text. The text is:
<answer>

AI: <response>

Human: Now, based on your assessment of the text’s style, create a JSON object where you score each of the five stylistic categories—excitable, stern, dramatic, rambling, and melancholy—on a scale from 0 to 10. Each category should receive a score reflecting its intensity in the text, with 0 indicating no presence and 10 indicating maximum presence. Ensure that your scores are represented accurately in the JSON, with each style as a key and the assigned score as its value.

AI: <response>

A.2.3. Settings

Each LLM assessed each answer ten times, with the scores averaged. Before averaging, each LLM provided 10 (questions) * 2 (answers [CEO and AI]) * 10 (repeats) = 200 assessments (except Claude 2.1, which refused to respond in 20 assessments). After averaging, each LLM yielded 10 (questions) * 2 (answers [CEO and AI]) = 20 scores for each of excitable, stern, dramatic, rambling, and melancholy.

Temperature is a setting that controls the randomness of the response. It was 0.5 for the first AI response (to allow the qualitative analysis to highlight different aspects) and 0 for the second AI response (to ensure the quantitative analysis returned its most confident response and in the correct JSON format).

A.2.4. Results

TABLE A8. Cronbach’s alpha for communication styles by LLM (Study 1)

LLM	Cronbach’s alpha
Claude 2.1	0.921
Claude 3 Haiku	0.904
Claude 3 Sonnet	0.951
Llama 2	0.849
Mixtral	0.892
GPT3.5	0.898
GPT4	0.833
GPT4 Turbo	0.919

Note. First, for each LLM and each communication style a Cronbach’s alpha is calculated before averaging assessments. Second, these Cronbach’s alphas are averaged by LLM and shown here.

TABLE A9. Cronbach’s alpha for communication styles across LLMs (Study 1)

Communication style	Cronbach’s alpha
excitable	0.947
stern	0.953
dramatic	0.952
rambling	0.778
melancholy	0.963

Note. The Cronbach’s alphas shown here are for the measures used (i.e., after averaging assessments by LLM).

TABLE A10. Scores of CEO and AI answers on five communication styles (Study 1)

Communication style	AI	CEO	<i>p</i> -value
excitable	4.23 (2.32)	2.61 (1.63)	< .001
stern	2.43 (1.35)	3.33 (1.63)	< .001
dramatic	2.52 (1.62)	1.54 (0.77)	< .001
rambling	1.14 (0.75)	1.00 (0.79)	.043
melancholy	0.82 (0.94)	0.89 (0.97)	.390

Note. The means are provided with the standard deviation in parentheses. The *p*-value is from a paired *t*-test with 79 degrees of freedom.

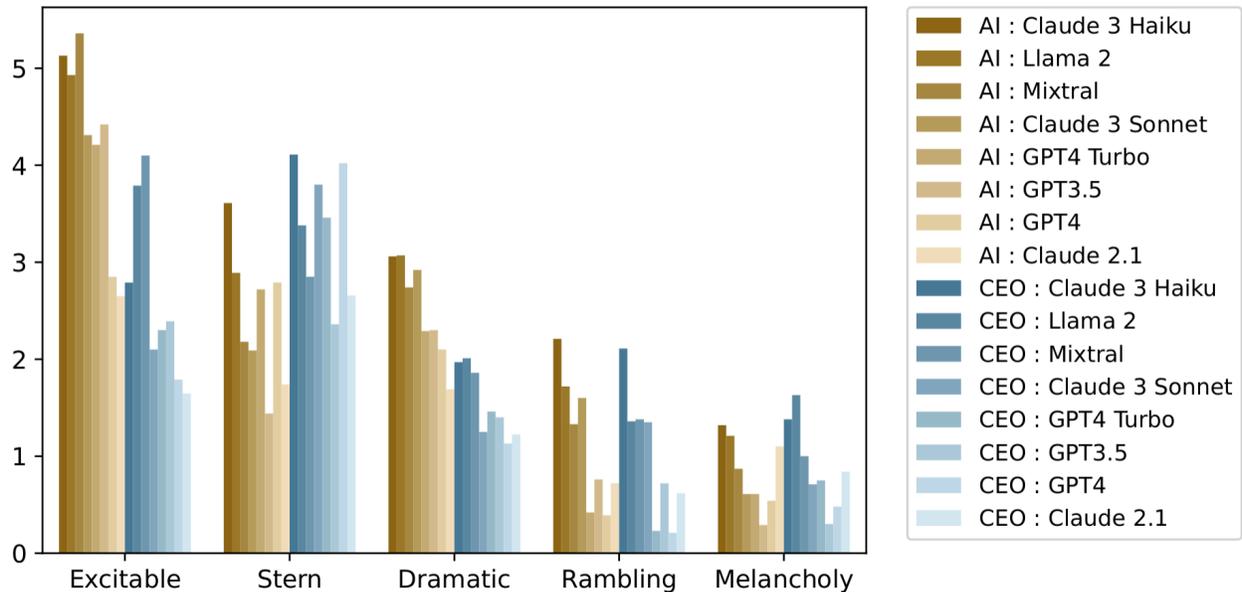


FIGURE A2. Average scores of CEO and AI answers on five communication styles by LLM (Study 1)

Note. Each bar represents the average of 10 answers assessed by 1 LLM.

A.3. Employees

We compared the sample with the organization overall for seniority level and tenure (see Table A11).

TABLE A11. A comparison of the sample with organization overall

		Sample	Organization	<i>p</i> -value
Senior	Senior management	13.0%	6.0%	0.016
	Other	87.0%	94.0%	
Tenure	Less than 1 year	13.0%	16.7%	0.208
	From 1 year to less than 3 years	44.0%	35.2%	
	3 years or more	43.0%	48.1%	
Sample size		100 employees	–	
Organization size		800+ employees	800+ employees	

Note. The sample percentages are from the time of the experiment (March 2024) and the organization percentages are from the time of a review (March 2025). The *p*-value is from a Chi²-test assuming 800 employees for the current organization.

We analyzed the accuracy of perceived origin (see Table A12) and perceived helpfulness (see Table A13) by employee subgroup.

TABLE A12. Accuracy of perceived origin by subgroup (Study 1)

Subgroup	Value	Number of employees	Mean	95% CI low	95% CI high
Senior management	0	87	0.583	0.544	0.622
Senior management	1	13	0.646	0.569	0.723
Tenure	$x < 3$	57	0.565	0.520	0.611
Tenure	$x \geq 3$	43	0.619	0.567	0.672

Note. Bootstrapped 95% CIs are shown, each based on 10,000 samples. Bootstrapping is at the employee level to take into account the nested data structure. Five employees did not provide their position and five did not provide their tenure.

TABLE A13. Regressions of perceived helpfulness by subgroup (Study 1)

	(1)	
	Perceived helpfulness	
	<i>b</i>	<i>se</i>
Perceived AI	−0.673	0.095
True AI	0.026	0.061
Perceived AI * Senior	0.008	0.202
Perceived AI * Tenure (≥ 3 years)	−0.239	0.133
Questions	FE	

Employees	FE	
Constant	4.033	0.045
<hr/>		
<i>n</i> observations	990	
<i>n</i> employees	99	
<i>F</i>	23.9	
<i>R</i> ²	0.572	
<hr/>		

Note. FE is fixed effects. Model 1 is a linear regression with standard errors clustered by employee and question. Six employees did not provide their position or tenure.

B. STUDY 2

B1. CEO and AI Answers

B1.1. Hershey

Financial analyst: What are your capacity expansion plans?

CEO: We've continued to invest in capacity in brands and businesses across the portfolio that have growth and opportunity ahead. Over the past couple of years, we've focused on Reese's—we were short on capacity—so that we could fulfill consumer demand. Then, the other big area of focus we had was on the Gummies side of the business, in Sweets, and the first part of this year, we complete and have capacity coming online for that business, that we will be able to leverage to better participate in that segment in the back half of the year.

AI: At Hershey, our capacity expansion plans are focused on several key areas to support our growth and meet increasing consumer demand. We are increasing our production capacity by 15% this year, particularly focusing on our Reese's brand and the Gummies segment within our sweets business. This expansion is crucial to fulfill consumer demand and leverage growth opportunities in these high-demand areas. Additionally, we have invested significantly in our technology and organizational capabilities to improve our end-to-end connectivity and efficiency. These investments include the implementation of the S/4 ERP system and enhancements in automation and process optimization, which will drive productivity across our supply chain and manufacturing operations. These strategic expansions and technological advancements will enable us to better meet consumer needs and support our long-term growth objectives.

B1.2. General Motors

Financial analyst: I'm interested in hearing sort of your updated views on competition from the Chinese outside China. What's GM's stance on this?

CEO: Yeah, it's a great question. And first of all, you know, I think, in general, we want to have our best products. And if there's a level playing field, then it's—you know, we want to compete based on product. I think you have to look at where is there a level playing field and—and what's happening around the world. But, you know, there's a lot that can happen from a regulatory or—you know, a trade perspective, but we're focused on is making sure we have great vehicles at the right price. So, what is going to help, you know, GM maintain its share around the world, you know, when you look at South America, the Chevy brand is incredibly strong. And we're going to continue to focus on having great designs with great—great product portfolio with the right features and functions. And we're constantly working on taking cost out of the system. So, it's a—there's value there as well. And that's the way we're going to compete around the world.

AI: GM views competition from Chinese automakers outside of China with a strategic and balanced approach. We recognize the increasing presence and capability of Chinese manufacturers, especially in regions like South America. While we advocate for fair competition, we believe that maintaining a level playing field is crucial. This means addressing both tariff and non-tariff barriers that can skew the market dynamics. Protectionism is not our preferred route; instead, we support a balanced trade policy that ensures equitable conditions for all players. By focusing on our strengths, such as innovation, quality, and customer experience, we aim to compete effectively in the global market without resorting to restrictive trade practices.

B1.3. Nvidia

Financial analyst: I wanted to follow up on the 40% of revenues coming from inference. That's a bigger number than I expected. Can you give us some sense of where that number was maybe a year before, how much you're seeing growth around LLMs from inference?

CEO: I'll go backwards. The estimate is probably understated. And—but we estimated it. And let me tell you why. Whenever—a year ago, the recommender systems that people are—when you run the internet, the news, the videos, the music, the products that are being recommended to you because as you know, the internet has trillions—I don't know how many trillions, but trillions of things out there and your phone is 3-inches square. And so the ability for them to fit all of that information down to something, such a small real estate, is through a system, an amazing system called recommender systems. These recommender systems used to be all based on CPU approaches. But the recent migration to deep learning and now generative AI has really put these recommender systems now directly into the path of GPU acceleration. It needs GPU acceleration for the embeddings. It needs GPU acceleration for the nearest neighbor search. It needs GPU acceleration for the re-ranking and it needs GPU acceleration to generate the augmented information for you. So GPUs are in every single step of a recommender system now. And as you know, recommender system is the single largest software engine on the planet. Almost every major company in the world has to run these large recommender systems. Whenever you use ChatGPT, it's being inferred. Whenever you hear about Midjourney and just the number of things that they're generating for consumers, when you when you see Getty, the work that we do with Getty and Firefly from Adobe. These are all generative models. The list goes on. And none of these, as I mentioned, existed a year ago, 100% new.

AI: Inference workloads now make up 40% of NVIDIA's revenue, a significant rise driven by the rapid adoption of large language models (LLMs) and generative AI technologies. A year ago, this percentage was notably lower, showing accelerated growth in AI deployment across sectors. This increase is fueled by the extensive use of our GPUs for both training and inference tasks. We measure inference contributions through sophisticated tracking of GPU utilization. Many GPUs serve dual purposes, but our detailed telemetry and analytics ensure robust metrics that differentiate between tasks, providing accurate revenue impact data. Generative AI, especially LLMs, is a key driver of this growth. These models have transformed data processing from retrieval-based to generative approaches, enabling applications like text-to-image and text-to-video. The transition to natural language queries has revolutionized data interaction. NVIDIA AI Foundations help businesses create custom LLMs and generative AI models with proprietary data, enhancing productivity and innovation across industries like healthcare, automotive, and finance. Collaborations with cloud providers like AWS, Google Cloud, and Microsoft Azure support this growth by offering scalable, AI-ready infrastructure. The surge in data center revenues driven by AI demand highlights the transformative impact of these technologies. Emerging AI factories, specialized data centers for AI processing, reflect the trend towards energy-efficient, high-performance computing solutions. In summary, the rise in inference-related revenues underscores the broad adoption of AI technologies, the effectiveness of our measurement systems, and the strategic expansion of our AI infrastructure.

B1.4. Amazon

Financial analyst: On logistics, in September, you launched Amazon Supply Chain. Can you just help us understand the opportunity you see there?

CEO: I think that it's interesting what's happening with the business we're building in third-party logistics. It really kind of, in some ways, mirrors some of the other businesses we've gotten involved in, AWS being an example of it, even though it's—they're very different businesses in that we realized that we had our own internal need to build a bunch of these capabilities, and we figured that there were probably others who had those same needs, and we decided to build services out of them. So as our business has grown, it turns out to be pretty hard work to actually import items from overseas, get them through customs and through the border and then ship them from that point to various facilities. And then it turns out that you don't want to store those facilities in fulfillment centers because that space is really scarce, so you'd like to have them in upstream storage facilities that are very inexpensive. And then you'd like to have a way to be able to know when your more scarce supply in the fulfillment centers needs replenishment and be able to do it automatically from those upstream storage facilities.

AI: Absolutely. When we look at Amazon Supply Chain, we see a significant opportunity to leverage our extensive logistics network and technology expertise. This initiative allows us to offer a comprehensive suite of services to businesses of all sizes, helping them manage their supply chain more efficiently. We're already well down the road in this journey, having built a robust logistics infrastructure that includes fulfillment centers, transportation, and last-mile delivery capabilities. These assets give us a strong foundation to extend our logistics services beyond our own needs to support other companies. In terms of CapEx, it's important to note that while there is an investment component, much of the infrastructure we need is already in place, thanks to years of building and refining our logistics network. Future investments will focus on enhancing our capabilities, incorporating more automation, and scaling up our services globally. Overall, we see this as a natural extension of our business that can drive growth, improve efficiency for our customers, and ultimately create more value across the board.

B2. Additional Analyses

TABLE B1. Regressions of perceived helpfulness and buying intentions (Study 2)

	(1)		(2)	
	Perceived helpfulness		Share purchase intentions	
	<i>b</i>	<i>se</i>	<i>b</i>	<i>se</i>
Labeled AI	-0.190	0.091	-0.192	0.090
True AI	0.835	0.091	0.636	0.090
Labeled AI * True AI	0.088	0.129	0.078	0.127
Companies	FE		FE	
Participants	RE		RE	
Constant	3.101	0.090	2.873	0.094
<i>n</i> observations	864		864	
<i>n</i> participants	216		216	
<i>Chi</i> ²	244.4		176.0	
<i>R</i> ²	0.191		0.125	

Note. FE is fixed effects and RE is random effects. Models 1 and 2 are linear regressions.

References

- Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. (2019). Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal*, *40*(11), 1705–1732.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*, 24824–24837.