

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

Working Paper

The Wade Test: Generative AI and CEO Communication

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 (2024) : The Wade Test: Generative AI and CEO Communication, CESifo Working Paper, No. 11316, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/305558>

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.

The Wade Test: Generative AI and CEO Communication

Prithwiraj Choudhury, Bart S. Vanneste, Amirhossein Zohrehvand

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

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

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

The Wade Test: Generative AI and CEO Communication

Abstract

Can generative artificial intelligence (AI) transform the role of the CEO by effectively automating CEO communication? This study investigates whether AI can mimic a human CEO and whether employees' perception of the communication's source matter. In a field experiment with a firm, we extend the idea of a Turing test (i.e., a computer mimicking a person), to the idea of generative AI mimicking a specific person, namely the CEO. We call this the "Wade test" and assess if employees can distinguish between communication from their CEO and communication generated by an AI trained on the CEO's prior communications. We find that AI responses are correctly identified 59% of the time, somewhat better than random chance. When employees believe a response is AI generated, regardless of its actual source, they perceive it as less helpful. To assess causal mechanisms, a second study with a general audience, using public statements from CEOs and from an AI intended to mimic those CEOs, finds that AI-labeled responses (irrespective of their actual source) are rated as less helpful. These findings highlight that, when using generative AI in CEO communication, people may inaccurately identify the source of communication and exhibit aversion towards communication they identify as being AI generated.

Prithwiraj (Raj) Choudhury
Harvard Business School, Harvard University
Cambridge / MA / USA
pchoudhury@hbs.edu

Bart S. Vanneste
UCL School of Management
University College London / United Kingdom
b.vanneste@ucl.ac.uk

Amirhossein Zohrehvand
Faculty of Science, Science Based Business
Leiden University / The Netherlands
a.h.zohrehvand@sbb.leidenuniv.nl

August 23, 2024

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.

1. Introduction

How generative artificial intelligence (AI) will affect firms and the future of work is a question of great interest to strategy scholars (e.g., Berg et al. 2023), policy makers (e.g., Cazzaniga et al. 2024), and managers (e.g., Microsoft 2024). This line of inquiry follows in the footsteps of studies on AI adoption within firms, its impact on individual performance (e.g., Allen and Choudhury 2022, Gaessler and Piezunka 2023, Jia et al. 2024), and its potential to substitute, complement, or alter jobs (e.g., Felten et al. 2021). Recent advances in generative AI, which can create high-quality data such as text, images, audio, or video, are expected to amplify AI's impact on organizations and their employees (Berg et al. 2023). Progress in large language models (LLMs) is particularly noteworthy, due to their ability to generate natural language, the primary mode of communication both within and outside organizations (Yates and Orlikowski 1992). For example, LLMs can aid communications internally by facilitating or reducing the need for meetings between geographically dispersed R&D engineers (Zhang et al. 2024) or externally by supporting or replacing customer service agents interacting with clients (Yang et al. 2024). This raises the question of whether generative AI could play an important role and even substitute humans in the act of communication across organizational roles (Eloundou et al. 2024).

For strategy scholarship, a central question emerging is whether and how effectively generative AI could transform the role of the CEO. The CEO is essential not only for formulating, but also for communicating firm strategy (e.g., Hambrick 2007). There is extensive literature on the importance of communication in performing the CEO's role (Bandiera et al. 2020, Kotter 1990). CEOs communicate both with external constituencies such as shareholders (through several devices, including letters published in the firm's annual report, Kaplan, 2008), and especially with internal constituencies or employees of the firm (e.g., Reyt and Wiesenfeld 2014).¹ Using fine-grained data collected in a CEO diary, Bandiera et al. (2020) document that CEOs spend on average 70% of their time interacting with others (either face to face via meetings or plant visits, or virtually via phone, videoconferences, or

¹ There is a vast literature highlighting the importance of intrafirm communication for CEOs, including notable work by organizational scholars such as Schein (1996), strategy scholars such as Van den Steen (2018), and economists such as Rotemberg and Saloner (2000).

emails). 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. Generative AI could play an important role in automating communication within organizations, including communication by the CEO, which could help free up time for CEOs to focus on other tasks. This leads us to ask, can generative AI effectively automate CEO communication? We further ask if employees can detect the source (generative AI or human CEO) of CEO communication and whether their perception of the communication's source affects judgements about the communication's helpfulness?

Our approach to comparing generative AI with a human CEO in performing CEO communication builds on the classic Turing test (Turing 1950). Originally called an imitation game, the Turing test is a blind test where a human guesses whether they are communicating with a human or a machine. If the human's guesses about the other's identity are not better than random chance, the machine has successfully passed the Turing test. Passing the Turing test has been a longstanding challenge. The Loebner Prize, an annual competition held from 1991 to 2019, recognized the top-performing chatbots each year, but none claimed the larger prizes for passing the Turing test (Neufeld and Finnestad 2020). With the recent advances in generative AI, especially large language models (LLMs), passing the Turing test now appears feasible (Biever 2023). For example, in an online experiment, 1.5 million users correctly identified whether the other was a human or chatbot only 68% of the time, with accuracy dropping to 60% when the other was a chatbot (Jannai et al. 2023).

In a field experiment with a firm, we extend the idea of a Turing test from mimicking a person to mimicking a *specific* person, namely the CEO. We call this the “Wade test,” named after the CEO. The machine is a generative AI trained by the firm on the CEO's prior communication. As summarized in detail later, the experiment involved four main steps: (1) crowdsourcing questions for the CEO from employees at the firm, (2) asking the human CEO, i.e., Wade, to respond to the crowdsourced questions, (3) creating answers to these questions using generative AI, and (4) setting up an experiment where employees at the firm were randomly exposed to either Wade's own answers or answers generated by the

AI and asking them to guess whether the human or AI had generated the answers and assess the helpfulness of the answers.

Using 105 (approximately 13% of) employees of the firm, we investigate if the machine passes the Wade test (i.e., whether its answers to 10 predefined questions are indistinguishable from those of the CEO) and whether employees perceive messages differently depending on their belief about whether they are from the CEO or the machine. We find 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%. In sum, generative AI nearly passes the Wade Test and the communication that it generates could be thought of as an authentic substitute to communication generated by the human CEO.

Given that our research question asks whether generative AI can effectively automate communication attributed to the CEO, we also test whether employees rated answers generated by the AI as equally helpful as the answers generated by the human CEO. We find that passing the Wade test is consequential. Our analysis suggests that answers generated by the machine and the CEO are rated as equally helpful. However, 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. In other words, the effectiveness of generative AI communication depends on whether employees are able to guess which content is generated by the CEO vis-a-vis generated by the generative AI algorithm. This finding suggests evidence of algorithm aversion (e.g., Dietvorst et al. 2015) of employees towards generative AI communication.

To assess causality for the second finding, we conduct an additional experiment with a general audience. 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). The researchers randomized whether answers were labeled as being generated by the human CEO or by generative AI. Importantly, regardless of their actual origin, answers labeled as AI generated were seen as less helpful than those labeled as generated by

the CEO. Thus, whether the message is perceived by receivers as generated by AI (study 1) or labeled by researchers as being generated by AI (study 2), a message's potency is reduced. This confirms that human subjects exhibit algorithm aversion towards generative AI communication.

This study contributes to research on how generative AI might impact the future of work (e.g., Felten et al. 2023) and takes an important strategy lens in exploring whether generative AI might effectively automate communication by CEOs. Generative AI may fundamentally change the way organizations communicate and operate, prompting research into how roles and the performance of tasks within organizations will change (Dell'Acqua et al. 2024, Eloundou et al. 2024; Otis et al. 2024). A vibrant literature has emerged investigating the impact of AI on employee roles, including consultants, customer support agents, and IT support specialists (Allen and Choudhury 2020, Brynjolfsson et al. 2023, Dell'Acqua et al. 2024, Jia et al. 2024). However, the CEO's role, perhaps the most pivotal in communicating firm strategy, has remained unaddressed until now. We begin this exploration by focusing on communication, a primary part of a CEO's role and a domain where LLMs demonstrate proficiency. Our approach is to extend the Turing test (human versus machine) to the Wade test (specific human versus machine). CEO communication is influential not only for its content but also because of who conveys it. Hence, we highlight that for generative AI to effectively transform the CEO role, it is important to consider not only the technical capabilities of generative AI but also the perceptions it creates among receivers.

2. Background

2.1. CEO communication

The CEO is the most influential person in the firm (Hambrick and Quigley 2014), responsible for formulating and communicating the firm's strategy (Hambrick and Mason 1984). CEO communication is thus important for an organization (Antonakis et al. 2022, Helfat and Peteraf 2015), also as evidenced by the fact that CEOs spend approximately 70% of their daily schedule on communication (Bandiera et al. 2020). Through communication, a CEO engages with both internal and external stakeholders to direct

their attention (Ocasio et al. 2018), coordinate resources, and drive strategic change (Helfat and Peteraf 2015). Recognizing the importance of CEO communication, scholars have documented and studied its various forms (Liu, EH et al. 2023), including earnings conference calls (Crilly 2017), shareholder letters (Fanelli et al. 2008), CEO letters to other top managers (Kaplan 2015), internal social media posts (Leonardi 2018), and emails (Reyt and Wiesenfeld 2014).

The evolving demands of effective communication have led CEOs to increasingly rely on others' expertise for crafting clear messages and reaching multiple audiences (Porter and Nohria 2018, Sadun et al. 2022). This development has been reinforced by the widespread adoption of communication technologies that amplify the reach of CEO communication (e.g., Heavey et al. 2020). For example, for external communication, this includes the crafting of annual letters to shareholders—a critical task that requires strategic insight and a well-structured narrative (Amernic and Craig 2007). Likewise, for internal communication, the creation and communication of vision and mission statements relies on collaborative efforts (e.g., Fitzsimmons et al. 2022). Thus, to reinforce the organization's strategy, CEOs turn to others for crafting and conveying their messages effectively.

2.2. AI-generated CEO communication

Generative AI, particularly LLMs, also has the potential to aid CEOs in communication. Generative AI provides models that can create new data, including text, images, sounds, and video. During training on large datasets, these models learn the distribution of the data (Murphy 2023). After training, they can draw from this distribution to create new data that is similar but not identical to the original data. A notable category of generative AI is LLMs, which generate text as output. This capability allows LLMs to execute a broad range of tasks that rely on natural language. Recent research has documented different organizational use cases of LLMs. For example, in customer interactions, LLMs are used to prescreen sales calls (Jia et al. 2024), function as customer chatbots (Gnewuch et al. 2023), and provide call feedback to human agents (Brynjolfsson et al. 2023, Tong et al. 2021). In interactions within organizations, they facilitate communication in teams that are geographically distant (Zhang et al. 2024). Furthermore, LLMs are integrated for standalone work, for example for advice drafting (Dell'Acqua et al.

2024, Otis et al. 2024), idea generation (Boussioux et al. 2024, Doshi and Hauser 2024), or decision making (Doshi et al. 2024).

Generative AI and LLMs have several appealing characteristics that may be useful in CEO communication. First, once trained, they can produce content at low cost (Musser 2023) compared to CEOs, who incur high cost of time and potentially cognitive effort for generating communication. Second, LLMs can produce content at scale and have the ability to tailor messages to specific contexts (Matz et al. 2024), which may benefit CEOs seeking to reach a broad audience (e.g., Heavey et al. 2020). Third, LLMs can create and maintain different personas (e.g., Tseng et al. 2024), enabling them to mimic the CEO's tone and voice. Since the CEO is the leader of the organization, audiences desire authenticity in their communication (Gardner et al. 2011).

2.3. Reception of AI-generated CEO communication

The use of generative AI in CEO communication could however change how an audience perceives and evaluates the communication. Even if AI-generated communication can mimic that of the CEO, the audience may perceive it differently. First, in general, people may prefer receiving communication from a human CEO than from an AI, based on the literature on algorithm aversion. Second, specifically in the context of CEO communication, it matters that the sender is not only a human (versus AI) but also a specific individual (namely the CEO).

Thus, we suggest that people judge messages differently depending on whether they perceive the source to be the CEO or AI for two main reasons. First, the literature on algorithm aversion, which is defined as “biased assessment of an algorithm which manifests in negative behaviors and attitudes towards the algorithm compared to a human agent” (Jussupow et al. 2020: 4), suggests that people prefer to rely on other humans than on AI. For example, Dietvorst et al. (2015) document that in the case of a prediction task, humans exhibit aversion towards algorithmic prediction and prefer to depend on human prediction instead. Other studies also report evidence of algorithm aversion (e.g., Allen and Choudhury 2022, Castelo et al. 2019, Dietvorst et al. 2018, Liu, M et al. 2023, Tong et al. 2021, but see also Dargnies et al. 2024, Logg et al. 2019). One reason is that people may perceive algorithms to be less capable and

knowledgeable than a human (Burton et al. 2020, Luo et al. 2019, Reich et al. 2021). As applied to CEO communication, a CEO is perceived as knowledgeable about the issues pertaining to the firm (Finkelstein et al. 2009), which does not necessarily hold for an AI. Therefore, people may perceive the AI-generated content to be inferior and less helpful than that being produced by the CEO.

Second, the CEO is not just any individual but the leader of the organization. Their communication is influential not only for its content but also because of who conveys it. As a result, messages coming from a CEO are received differently than those from any other source. The position of the CEO lends credibility, legitimacy, and symbolism to the communication (Hambrick and Lovelace 2018, Men 2012, Suchman 1995). A credible source is a source that the audience perceives as having the ability and willingness to produce correct content (Hovland and Weiss 1951, Pornpitakpan 2006). The credibility of a CEO's communication is partly based on their ability to fulfill the promises made in their messages, such as committing to a strategy that the organization will adopt. Legitimacy is "a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions" (Suchman 1995, 574). The legitimacy of a CEO's communication is largely derived from their position, which inherently carries a sense of authority and appropriateness (Finkelstein et al. 2009). Finally, a CEO's communication may hold symbolic significance, meaning that a message conveys information beyond its literal content (Hambrick and Lovelace 2018). Communication that is high on credibility, legitimacy, and symbolism is perceived as more useful and persuasive (Bolsen et al. 2019, Hambrick and Lovelace 2018, Pornpitakpan 2006, Tyler 2006). To the extent that AI-generated communication has less credibility, legitimacy, or symbolism, we expect that it will be considered less helpful than communication attributed to the CEO.

Figure 1 summarizes the relationships we investigate. First, we assess the extent to which recipients can identify the source of communication, whether it is from a CEO or an AI (relationship 1: from True AI to Perceived AI). Second, we investigate whether perceptions about a communication's source affect its perceived helpfulness (relationship 2: from Perceived AI to Perceived helpfulness), while

accounting for possible differences between CEO and AI communication (relationship 3: from True AI to Perceived helpfulness).

[[INSERT FIGURE 1 ABOUT HERE]]

3. Study 1

We conducted a pre-registered field experiment with 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 impacts its perceived helpfulness.²

3.1. Context

Pulse is a technology company with more than 800 employees. The CEO co-founded the company and is well known within the company. The CEO was motivated to engage in this study for two reasons. First, the CEO of Pulse, like other CEOs, is shadowed by executive assistants who often respond to communication on behalf of the CEO. The CEO was intrigued by the prospect of automating this process. Second, leading a technology company, the CEO wanted to see if the LLM could be a test case for LLM models in the future.

The company created the CEO bot in three steps. 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. In the first step, all available textual communication written by the CEO from internal and external sources (such as emails, Slack channels, and documents) was collected. The goal was to build a mapping between the CEO's style and that of a "typical" manager. For this purpose, the CEO's text was made available to an LLM, which converted sentences written by the CEO into sentences that might be written by a typical manager. As an example, if the CEO in question had some favorite phrases or words (e.g., "basically" or "thanks a million"), the mapping removed these phrases and words and generated the text in the style of an average manager. In this way, the first step yielded a mapping of language actually written by the CEO into language that was representative of an average manager writing text in English (i.e., CEO text to

² The pre-registration is available here: https://aspredicted.org/NB1_F42

plain text). This step was critical to create the “rephraser,” which maps text in the opposite direction (i.e., translating plain English text to text that the CEO could plausibly write), in the third step (explained below). In the second step, all available company documents (e.g., Slack data, emails, and company reports) were made available to a different LLM, such that it could answer questions pertaining to the policies and future directions of the company. It is important to note that the answers generated by this LLM were written in plain text.

In the third and crucial step, a “rephraser” algorithm was written to rephrase the LLM’s answers from the second step to answers that could conceivably have been written by the CEO in question. This entailed flipping the mapping generated in Step 1, so that the mapping was now from typical manager language to the CEO’s language (plain text to CEO text). Since Step 3 reverses the mapping generated in Step 1, it was called the “rephraser” step. This step was crucial to “rephrase” the content generated in Step 2 into the communication style of the CEO in question. At the end of this step, the algorithm was able to generate an answer to any question, answered by the LLM using company documents, rephrased in the style typically used by the CEO.

3.2. Methods

3.2.1. Participants

All Pulse employees were invited to participate in the study, and 105 did. Participants included senior management (13%) and other (87%). Participants had worked at the company less than one year (13%), from one year to less than three years (44%), or three years or more (43%). Hence, the typical participant has been exposed to the CEO’s communication over a sustained period.

3.2.2. Procedure

Similar to in 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 recently submitted in an Ask-Me-Anything (AMA) forum. From this pool, an LLM (not the one trained to mimic the CEO) extracted 10 themes and one exemplar question per theme. We used the 10

resulting questions (see Table 1). The CEO and AI separately responded to the questions. All answers were retained unchanged.

[[INSERT TABLE 1 ABOUT HERE]]

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.³ 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).⁴

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 provides 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 provides 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. We include fixed

³ 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.

⁴ 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 for questions and random effects for employees, which is more efficient than fixed 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.⁵

[[INSERT TABLE 3 ABOUT HERE]]

Model 2 of Table 3 provides an ordinary least squares (OLS) regression with dependent variable perceived helpfulness ($M = 3.621$, $SD = 1.005$). We include fixed effects for questions and for employees, because perceived AI is not orthogonal to employees. 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 ($b = 0.020$; $p = .625$). 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.⁶

3.4. Additional analysis

For details, see the Online Appendix.

3.4.1. 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 generative AI (for details, see

⁵ This is $E[P(Y = 1 \mid \text{true AI} = 1)] - E[P(Y = 1 \mid \text{true AI} = 0)]$ and calculated as $1/J \sum_j E[P(Y = 1 \mid \text{true AI} = 1, \text{question} = j)] - 1/J \sum_j E[P(Y = 1 \mid \text{true AI} = 0, \text{question} = j)]$.

⁶ 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).

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 3 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 3 ABOUT HERE]]

3.4.2. Differences between questions

We investigated if accuracy of perceived origin differed by question specificity, which is the degree to which the question uniquely applies to the focal company and not to other companies. If the question is specific, then the AI may struggle to provide a good answer, allowing employees to distinguish it from the CEO's answer.

We used a generative AI approach similar to the one discussed above. We asked the same eight LLMs to assess the specificity of each question, first qualitatively and then quantitatively using a scale from 0 (question is entirely generic and applicable to many companies) to 10 (question is highly specific and applicable only to the focal company). Each LLM assessed each question 10 times, with the scores averaged. The resulting scores had low reliability between LLMs with a Cronbach's alpha of .629. We investigated the reliability within each LLM and found that three older or smaller LLMs (Claude 2.1, Claude 3 Haiku, GPT3.5) had a Cronbach's alpha of less than .7, a typical lower threshold (Nunnally

1978). The results presented here are based on the remaining five LLMs. The Online Appendix presents the results for all eight LLMs, which are similar.

Figure 4 shows the specificity and accuracy for each question. They have a correlation of .657. Thus, CEO and AI answers are more distinguishable if they relate to questions more unique to the focal company. This finding is consistent with the explanation that AI may find answering company-specific questions more challenging (e.g., because the relevant information was not part of its training data). As further corroboration, we included the questions in Table 1 but had to omit some text for four questions to preserve anonymity. Only these four questions have accuracy with confidence intervals above 50%.

[[INSERT FIGURE 4 ABOUT HERE]]

3.4.3. Differences between employees

We analyzed the accuracy of perceived origin by subgroup of employees. 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 ($b = .008, p = .957$) but is stronger for employees with longer tenure ($b = -.239, p = .011$). Thus, employees who joined earlier find answers perceived to be from the AI even less helpful than those who joined later.

4. Study 2

Study 1 revealed that employees regarded messages they believed came from the CEO as more helpful than those they thought were generated by AI. We cannot interpret that finding causally, because

employees may have attributed more helpful answers to the CEO. In the pre-registered study 2, we manipulate a message's perceived origin to provide causal evidence for the effect of a message's perceived origin on its perceived helpfulness.⁷

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 criteria.⁸ We opened 252 positions and paid a fixed fee of \$2.05. We excluded 21 participants whose completion took less than two minutes or more than 15 minutes, and 13 participants who failed an attention check, yielding a sample size of 218 (56% female; M age = 34.5; SD age = 10.5).

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. 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 GPT4o 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 $872 = 218 \text{ (participants)} * 4 \text{ (questions)}$.

4.1.3. Measures

⁷ The pre-registration is available here: https://aspredicted.org/D2Z_52R

⁸ 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).

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).

4.2. Results

Model 1 of Table 4 provides an OLS regression with dependent variable perceived helpfulness ($M = 3.288$, $SD = 1.177$), 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 ($b = -0.215$; $p = .002$), which gives a standardized mean difference of -0.182 . 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 ($b = 0.388$; $p < .001$). Answers actually from the AI were perceived as more helpful. The findings for share purchase intention ($M = 2.853$, $SD = 1.223$) are similar (see model 2). The AI-labeled answers reduce share purchase intentions ($b = -0.204$; $p = .002$), whereas actual AI answers increase them ($b = 0.290$; $p < .001$). Hence, labeling answers as from an AI reduces their potency, even as actual AI answers compare favorably to those of the CEO.

[[INSERT TABLE 4 ABOUT HERE]]

4.3. 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 (65.1%, $\text{Chi}^2(1) = 25.3$, $p < .001$). Second, we assessed whether the effect of labeled AI differed for actual CEO and AI answers. The effect of labeled AI is more negative for actual AI answers, but not significantly so (perceived helpfulness: $b = -0.136$, $p = 0.314$; share purchase intention: $b = -0.060$, $p = 0.642$).

5. Discussion

In this study, we investigate generative AI's ability to mimic CEOs' communication and its effect on recipients. First, we conducted a field experiment with an AI that replicates a CEO's communication style and assessed employees' perceptions of the AI-generated content. In what we call the Wade test, we assessed whether employees could distinguish between messages generated by the CEO in question and those generated by AI, as well as their perceptions of the messages based on the perceived source. The findings indicate that employees were only slightly better than chance at distinguishing between AI-generated and CEO-generated messages. Moreover, messages perceived as AI generated were rated as less helpful. Second, in an online experiment, we manipulated the source label and investigated its effect on a message's perceived helpfulness. Again, messages labeled as from the AI were seen as less helpful. These results suggest that, while generative AI can support CEO communication, its effectiveness is potentially hindered if identified as AI generated.

Our findings have implications for several streams of research. First, we contribute to the literature on the impact of AI on organizations and the future of work (Allen and Choudhury 2022, Berg et al. 2023, Dell'Acqua et al. 2024, Felten et al. 2021, Jia et al. 2024). Building on existing research that has examined how AI impacts employees' performance in areas such as preparing advice (Dell'Acqua et al. 2024) or interacting with customers (Brynjolfsson et al. 2023), we shift the focus to AI's impact on the CEO role, primarily communication. Given the significance of this task for CEOs and for communicating firm strategy (Bandiera et al. 2020), we explore the potential for generative AI to complement or even substitute for this task. An important condition for augmenting a task with AI is for it to be able to perform the task well enough (Dell'Acqua et al. 2024), a condition that applies to CEO communication too. We extended the classic Turing test (Turing 1950) on human versus machine by proposing the Wade test on *specific* human versus machine. In this study, the specific human is the CEO of a firm. We provide evidence that generative AI can produce communication that employees have difficulty distinguishing from their CEO's communication, and yet document evidence of algorithm aversion towards

communication perceived to be generated by AI. Thus, we highlight both the potential and limitations of generative AI to impact job design and the future of work at the executive level.

Second, we contribute to the literature on the organizational use of algorithms by highlighting how algorithm aversion extends to the realm of generative AI. The algorithm aversion literature has thus far studied the use of algorithms as a prediction and decision support tool (as opposed to a content generation tool) and contends that people evaluate content quality based on whether they think it was created by an algorithm or a human (Dietvorst et al. 2015, 2018). We extend this literature by comparing an algorithm designed to generate content and a specific human. In our context, the specific human is the CEO. For the position of the CEO, what matters is not only what is communicated but also who communicates it. We show that people find the content generated by an AI that mimics a specific person less helpful.

Third, we contribute to the literature on the role of AI in CEO communication and strategic decision making (e.g., Choudhury et al. 2019; Csaszar et al. 2024; Doshi et al. 2024) by offering a novel perspective on CEO involvement in the communication process. Traditionally, the literature has studied messages originating from the CEO or where the CEO has been involved in their creation (Liu et al. 2023a). Because of the close connection between communication and the individual, extensive research has used CEO communication to infer personal attributes about the CEO such as personality traits (Harrison et al. 2019) or cognitive flexibility (Kiss et al. 2020). We shift the focus to communication that can be generated without direct CEO involvement. This shift is important because, when removing or relaxing the constraint on CEO involvement, CEO communication is no longer bounded and limited by CEOs' cognitive capabilities (Helfat and Peteraf 2015) but becomes scalable. Our insights are also relevant to the emerging literature on the use of AI in strategic decision making. In this literature, Doshi et al. (2024: 8) argue that when "language models become large, abilities emerge that are absent in smaller models, including answering previously unseen questions, performing arithmetic, and reasoning over multiple steps." This study documents evidence of generative AI "answering questions" on behalf of the actual CEO.

Future research can build on this study to explore several areas regarding generative AI in CEO communication. The first area is examining different modes of AI-generated communication. We focused on written text. Other possibilities include speech or video and in fact, recent examples include leaders using AI technologies to remotely “give” speeches (Zhuang 2024). Second, whereas we studied algorithms that sought to closely match the CEO’s style, generative AI can adopt various styles, deviating from the primary communication style of a CEO. Future research could explore how changes in the style of AI communication influence message recipients’ sense of the authenticity and success of the communication. Future research could also focus on algorithm aversion towards generative AI content attributed to the CEO among external stakeholders such as analysts, shareholders of the firm, customers, and policy makers.

6. Conclusion

This study explores the accuracy and effectiveness of using generative AI to replicate CEO communication. We found that generative AI could mostly pass as authentic CEO communication; employees correctly identified the source only slightly better than chance would be. This finding suggests a potential for AI to effectively support CEOs in their communication. However, employees rated responses as less helpful when they believed them to be from the AI, underscoring the importance of perception in evaluating AI content. Our findings contribute to the ongoing discussion on the role of AI in organizational settings, highlighting both the promise of AI in augmenting strategic tasks such as communication by the CEO and the challenges. In sum, the potential of generative AI to augment and transform key roles such as that of the CEO depends on careful consideration of human perceptions towards algorithmic content.

References

Allen R, Choudhury P (2022) Algorithm-augmented work and domain experience: The countervailing forces of ability and aversion. *Organization Science* 33(1):149–169.

- Amernic JH, Craig RJ (2007) Guidelines for CEO-speak: Editing the language of corporate leadership. *Strategy & leadership* 35(3):25–31.
- Antonakis J, d’Adda G, Weber RA, Zehnder C (2022) “Just Words? Just Speeches?” On the economic value of charismatic leadership. *Management Science* 68(9):6355–6381.
- Bandiera O, Prat A, Hansen S, Sadun R (2020) CEO behavior and firm performance. *Journal of Political Economy* 128(4):1325–1369.
- Berg JM, Raj M, Seamans R (2023) Capturing value from artificial intelligence. *Academy of Management Discoveries* 9(4):424–428.
- Biever C (2023) ChatGPT broke the Turing test—the race is on for new ways to assess AI. *Nature* 619(7971):686–689.
- Bolsen T, Palm R, Kingsland JT (2019) The impact of message source on the effectiveness of communications about climate change. *Science Communication* 41(4):464–487.
- Boussiou LN, Lane J, Zhang M, Jacimovic V, Lakhani KR (2024) Generative AI and creative problem solving. *SSRN Working Paper* 4533642.
- Brynjolfsson E, Li D, Raymond LR (2023) Generative AI at work. *NBER Working Paper* 31161.
- Burton JW, Stein MK, Jensen TB (2020) A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making* 33(2):220–239.
- Castelo N, Bos MW, Lehmann DR (2019) Task-dependent algorithm aversion. *Journal of Marketing Research* 56(5):809–825.
- Cazzaniga M, Jaumotte MF, Li L, Melina MG, Panton AJ, Pizzinelli C, Rockall EJ, Tavares MMM (2024) Gen-AI: Artificial intelligence and the future of work (International Monetary Fund). Retrieved August 21, <https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379>.
- Choudhury P, Wang D, Carlson NA, Khanna T (2019) Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal* 40(11):1705–1732.
- Cohen J (1992) A power primer. *Psychological Bulletin* 112(1):155–159.
- Crilly D (2017) Time and space in strategy discourse: Implications for intertemporal choice. *Strategic Management Journal* 38(12):2370–2389.
- Csaszar FA, Ketkar H, Kim H (2024) Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. *SSRN Working Paper* 4913363.
- Dargnies MP, Hakimov R, Kübler D (2024) Aversion to hiring algorithms: Transparency, gender profiling, and self-confidence. *Management Science*. Forthcoming.
- Dell’Acqua F, McFowland E, Mollick ER, Lifshitz-Assaf H, Kellogg K, Rajendran S, Kraye L, Candelon F, Lakhani KR (2024) Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* 24-013.

- Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology* 144(1):114–126.
- Dietvorst BJ, Simmons JP, 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 AR, Bell JJ, Mirzayev E, Vanneste B (2024) Generative artificial intelligence and evaluating strategic decisions. *SSRN Working Paper* 4714776.
- Doshi AR, Hauser OP (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.
- Fanelli A, Misangyi VF, Tosi HL (2008) In charisma we trust: The effects of CEO charismatic visions on securities analysts. *Organization Science* 20(6):1011–1033.
- 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.
- Finkelstein S, Hambrick DC, Cannella AA (2009) *Strategic leadership: Theory and research on executives, top management teams, and boards* (Oxford University Press, New York, NY).
- Fitzsimmons AB, Qin YS, Heffron ER (2022) Purpose vs mission vs vision: Persuasive appeals and components in corporate statements. *Journal of Communication Management* 26(2):207–219.
- Gaessler F, Piezunka H (2023) Training with AI: Evidence from chess computers. *Strategic Management Journal* 44(11):2724–2750.
- Gardner WL, Coglisier CC, Davis KM, Dickens MP (2011) Authentic leadership: A review of the literature and research agenda. *Leadership Quarterly* 22(6):1120–1145.
- Gnewuch U, Morana S, Hinz O, Kellner R, Maedche A (2023) More than a bot? The impact of disclosing human involvement on customer interactions with hybrid service agents. *Information Systems Research*. Forthcoming.
- Hambrick DC (2007) Upper echelons theory: An update. *Academy of Management Review* 32(2):334–343.
- Hambrick DC, Lovelace JB (2018) The role of executive symbolism in advancing new strategic themes in organizations: A social influence perspective. *Academy of Management Review* 43(1):110–131.
- Hambrick DC, Mason PA (1984) Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review* 9(2):193–206.
- Hambrick DC, Quigley TJ (2014) Toward more accurate contextualization of the CEO effect on firm performance. *Strategic Management Journal* 35(4):473–491.
- Harrison JS, Thurgood GR, Boivie S, Pfarrer MD (2019) Measuring CEO personality: Developing, validating, and testing a linguistic tool. *Strategic Management Journal* 40(8):1316–1330.

- Heavey C, Simsek Z, Kyprianou C, Risius M (2020) How do strategic leaders engage with social media? A theoretical framework for research and practice. *Strategic Management Journal* 41(8):1490–1527.
- Helfat CE, Peteraf MA (2015) Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic Management Journal* 36(6):831–850.
- Hovland CI, Weiss W (1951) The influence of source credibility on communication effectiveness. *Public Opinion Quarterly* 15(4):635–650.
- Imai K, Keele L, Tingley D (2010) A general approach to causal mediation analysis. *Psychological Methods* 15(4):309–334.
- Jannai D, Meron A, Lenz B, Levine Y, Shoham Y (2023) Human or not? A gamified approach to the Turing test. *arXiv Working Paper* 2305.20010.
- 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.
- Jussupow E, Benbasat I, Heinzl A (2020) Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. *European Conference on Information Systems* 28:1–16.
- Kaplan S (2008) Cognition, capabilities, and incentives: Assessing firm response to the fiber-optic revolution. *Academy of Management Journal* 51(4):672–695.
- Kaplan S (2015) Truce breaking and remaking: The CEO's role in changing organizational routines. *Cognition and Strategy*. Advances in Strategic Management (Emerald Group Publishing Limited, United Kingdom) 35:1–45.
- Kiss AN, Libaers D, Barr PS, Wang T, Zachary MA (2020) CEO cognitive flexibility, information search, and organizational ambidexterity. *Strategic Management Journal* 41(12):2200–2233.
- Kotter JP (1990) What leaders really do. *Harvard Business Review* 68(3):103–111.
- Leonardi PM (2018) Social media and the development of shared cognition: The roles of network expansion, content integration, and triggered recalling. *Organization Science* 29(4):547–568.
- Liu EH, Chambers CR, Moore C (2023) Fifty years of research on leader communication: What we know and where we are going. *Leadership Quarterly* 34(6):101734.
- Liu M, Tang X, Xia S, Zhang S, Zhu Y, Meng Q (2023) Algorithm aversion: Evidence from ridesharing drivers. *Management Science*. Forthcoming.
- Logg JM, Minson JA, Moore DA (2019) Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151:90–103.
- Luo X, Tong S, Fang Z, Qu Z (2019) Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science* 38(6):937–947.
- Matz SC, Teeny JD, Vaid SS, Peters H, Harari GM, Cerf M (2024) The potential of generative AI for personalized persuasion at scale. *Scientific Reports* 14(1):4692.
- McCracken G (1989) Who is the celebrity endorser? Cultural foundations of the endorsement process. *Journal of Consumer Research* 16(3):310–321.

- Microsoft (2024) 2024 Annual Work Trend Index. Retrieved August 22, 2024, <https://www.microsoft.com/en-us/worklab/work-trend-index/ai-at-work-is-here-now-comes-the-hard-part>
- Morwitz V (2014) Consumers' purchase intentions and their behavior. *Foundations and Trends in Marketing* 7(3):181–230.
- Murphy KP (2023) *Probabilistic machine learning: Advanced topics* (MIT press, Cambridge, MA).
- Musser M (2023) A cost analysis of generative language models and inference operations. *arXiv Working Paper* 2308.0374.
- Neufeld E, Finnstad S (2020) In defense of the Turing test. *AI & Society* 35(4):819–827.
- Nunnally JC (1978) *Psychometric theory* 2nd ed. (McGraw-Hill, New York, NY).
- Ocasio W, Joseph J (2018) The attention-based view of great strategies. *Strategy Science* 3(1):289–294.
- Otis N, Clarke R, Delecourt S, Holtz D, Koning R (2024) The uneven impact of generative AI on entrepreneurial performance. *SSRN Working Paper* 4671369.
- Porter ME, Nohria N (2018) How CEOs manage time. *Harvard Business Review* 96(4):42–51.
- Reich T, Kaju A, Maglio SJ (2023) How to overcome algorithm aversion: Learning from mistakes. *Journal of Consumer Psychology* 33(2):285–302.
- Reyt JN, Wiesenfeld BM (2014) Seeing the forest for the trees: Exploratory learning, mobile technology, and knowledge workers' role integration behaviors. *Academy of Management Journal* 58(3):739–762.
- Rotemberg JJ, Saloner G (2000) Visionaries, managers, and strategic direction. *RAND Journal of Economics* 31(4):693–716.
- Sadun R, Fuller J, Hansen S, Neal PJ (2022) The c-suite skills that matter most. *Harvard Business Review* 100(4):42–50.
- Schein EH (1996) Culture: The missing concept in organization studies. *Administrative Science Quarterly* 41(2):229–240.
- Suchman MC (1995) Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review* 20(3):571–610.
- 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.
- Tseng YM, Huang YC, Hsiao TY, Chen WL, Huang CW, Meng Y, Chen YN (2024) Two tales of persona in LLMs: A survey of role-playing and personalization. *arXiv Working Paper* 2406.01171.
- Turing AM (1950) Computing machinery and intelligence. *Mind* 59(236):433–460.
- Van den Steen E (2018) Strategy and the strategist: How it matters who develops the strategy. *Management Science* 64(10):4533–4551.

- Van Knippenberg D, Van Knippenberg B, De Cremer D, Hogg MA (2004) Leadership, self, and identity: A review and research agenda. *Leadership Quarterly* 15(6):825–856.
- Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, Chi E, Le QV, Zhou D (2022) Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35:24824–24837.
- Yang Z, Xu X, Yao B, Zhang S, Rogers E, Intille S, Shara N, Gao GG, Wang D (2024) Talk2Care: Facilitating asynchronous patient-provider communication with large-language-model. *arXiv Working Paper* 2309.09357.
- Yates J, Orlikowski WJ (1992) Genres of organizational communication: A structurational approach to studying communication and media. *Academy of Management Review* 17(2):299–326.
- Yin Y, Jia N, Waksalak CJ (2024) AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences* 121(14):e2319112121.
- Zhuang Y (2024) Imran khan’s ‘Victory Speech’ from jail shows A.I.’s peril and promise. *New York Times*. Retrieved August 22, 2024, <https://www.nytimes.com/2024/02/11/world/asia/imran-khan-artificial-intelligence-pakistan.html>.
- Zhang T, Koutsoumpis A, Oostrom JK, Holtrop D, Ghassemi S, de Vries RE (2024) Can large language models assess personality from asynchronous video interviews? A comprehensive evaluation of validity, reliability, fairness, and rating patterns. *IEEE Transactions on Affective Computing*. Forthcoming.

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)		(2)	
	Perceived AI		Perceived helpfulness	
	<i>b</i>	<i>se</i>	<i>b</i>	<i>se</i>
Perceived AI			−0.764	0.045***
True AI	0.824	0.136***	0.020	0.042
Question 1	0.322	0.218	4.056	0.246***
Question 2	−0.197	0.211	3.816	0.243***
Question 3	−0.372	0.212	3.948	0.238***
Question 4	−0.757	0.215***	4.229	0.238***
Question 5	−2.380	0.304***	4.219	0.238***
Question 6	−0.181	0.210	4.245	0.239***
Question 7	0.815	0.238***	4.053	0.243***
Question 8	−1.034	0.221***	4.054	0.239***
Question 9	−0.426	0.205*	4.058	0.238***
Question 10	−1.144	0.223***	4.040	0.237***
Employees	RE		FE	
<i>n</i> observations	1,050		1,050	
<i>n</i> employees	105		105	
<i>LL</i>	−640.8			
<i>F</i>			282.2***	
<i>R</i> ²			0.972	

Note. ***: $p < .001$, **: $p < .01$, *: $p < .05$. RE is random effects and FE is fixed effects. Model 1 is a logistic regression. Model 2 is an OLS regression.

Table 4. 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	−0.215	0.068**	−0.204	0.064**
True AI	0.388	0.068***	0.289	0.064***
Company: Hershey	3.440	0.089***	2.888	0.091***
Company: General Motors	2.681	0.089***	2.271	0.091***
Company: Nvidia	3.243	0.089***	3.022	0.091***
Company: Amazon	3.440	0.090***	3.060	0.092***
Participants	RE		RE	
<i>n</i> observations	872		872	
<i>n</i> participants	218		218	
<i>LL</i>	−1318.4		−1323.2	

Note. ***: $p < .001$, **: $p < .01$, *: $p < .05$. RE is random effects. Models 1 and 2 are OLS regressions.

Figure 1. Summary of key relationships

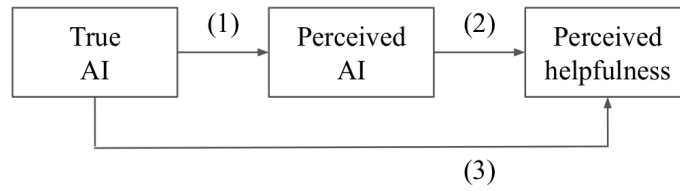
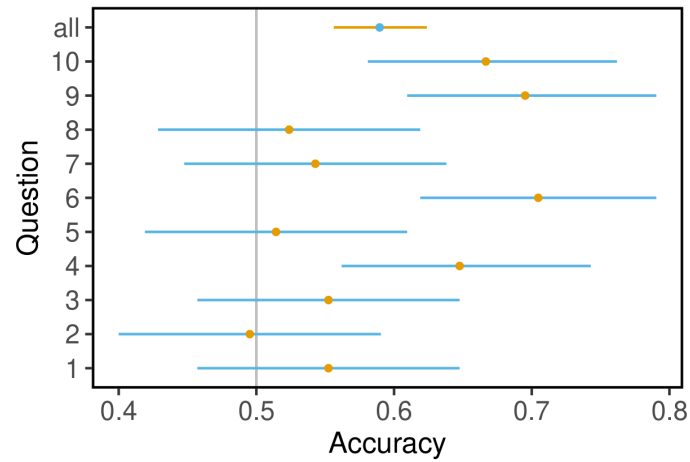
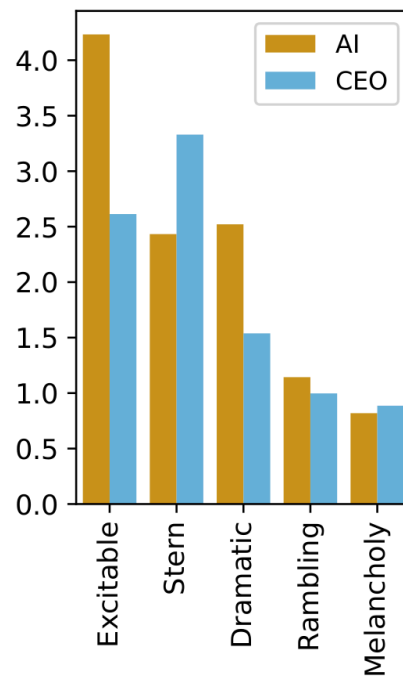


Figure 2. Accuracy of perceived origin (Study 1)



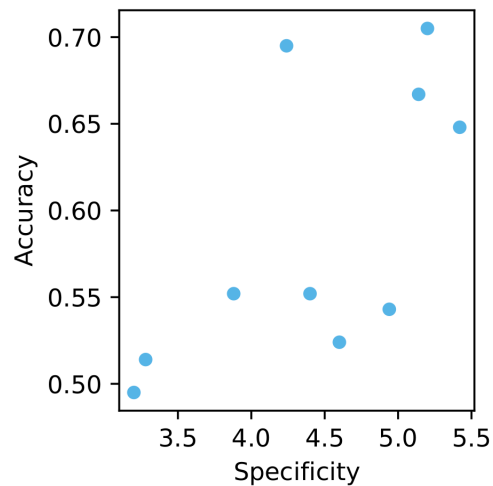
Note. Bootstrapped 95% CIs are shown, each based on 10,000 samples. For all questions combined (“all”), bootstrapping is at the employee level to take into account the nested data structure.

Figure 3. 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.

Figure 4. Question specificity and accuracy per question (Study 1)



Note. The correlation is .657. The specificity score is from 5 LLMs.

Online Appendix

A. Study 1

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.

A.1. Answers

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.

A.1.1. Large language models (LLMs)

We employed the eight LLMs shown in Table A2.

Table A2. 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.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 <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.1.3. Settings

Each LLM assessed each answer ten times, with the scores averaged. Before averaging, each LLM provided $10 \text{ (questions)} * 2 \text{ (answers [CEO and AI])} * 10 \text{ (repeats)} = 200$ assessments (except Claude 2.1, which refused to respond in 20 assessments). After averaging, each LLM yielded $10 \text{ (questions)} * 2 \text{ (answers [CEO and AI])} = 20$ 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.1.4. Results

Table A3. 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 A4. 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

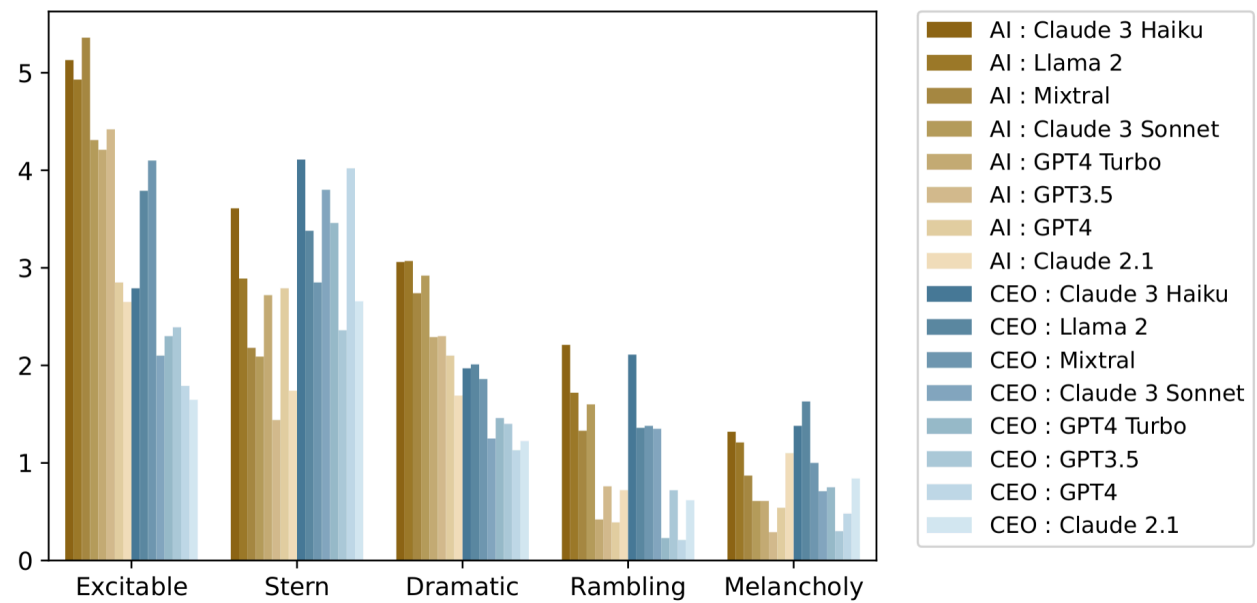
Note. The Cronbach's alphas shown here are for the measures used (i.e., after averaging assessments by LLM).

Table A5. 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 A1. 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.2. Questions

We used generative AI to assess the specificity of the questions to the focal company, using a similar approach as for assessing communication styles of the answers.

A.2.1. Large language models (LLMs)

We employed the same eight LLMs (see Table A2).

A.2.2. Prompts

The dynamic text <question> refers to the questions that the CEO and AI answered. To assess the specificity of the questions to the focal company, we changed the company's name to simply "company".

System: As a social scientist specializing in textual analysis, your task is to analyze the specificity of a question related to an anonymous company and then quantify this analysis.

Human: Carefully read the question provided below. Your task is to qualitatively assess the specificity of this question to the focal company. Reflect on the degree to which the question could be relevant to other companies or if it uniquely applies to the focal company. Indicate if this question, with the company's name changed accordingly, applies universally across all industries, selectively to particular industries, exclusively to the focal industry, or is only relevant to the focal company. You don't need to provide your assessment in a numerical form yet; just form your opinion on the level of specificity of the question to the company. The question is: <question>

AI: <response>

Human: Based on your qualitative assessment of the question's specificity to the focal company, now quantitatively score the specificity on a scale from 0 to 10. A score of 0 indicates that the question is entirely generic and applicable to many companies. A score of 10 signifies that the question is highly

specific and applicable only to the focal company. 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.2.3. Settings

Each LLM assessed each question ten times, with the scores averaged. Before averaging, each LLM provided 10 (questions) * 10 (repeats) = 100 assessments (and no refusals). After averaging, each LLM yielded 10 scores, one for each question.

Temperature was 0.5 for the first AI response and 0 for the second AI response.

A.2.4. Results

Table A6. Cronbach’s alpha for question specificity by LLM (Study 1)

LLM	Cronbach’s alpha
Claude 2.1	0.364
Claude 3 Haiku	0.562
Claude 3 Sonnet	0.928
Llama 2	0.747
Mixtral	0.770
GPT3.5	0.476
GPT4	0.945
GPT4 Turbo	0.970

Note. For each LLM, a Cronbach’s alpha is calculated before averaging assessments.

Table A7. Cronbach’s alpha for question specificity across LLMs (Study 1)

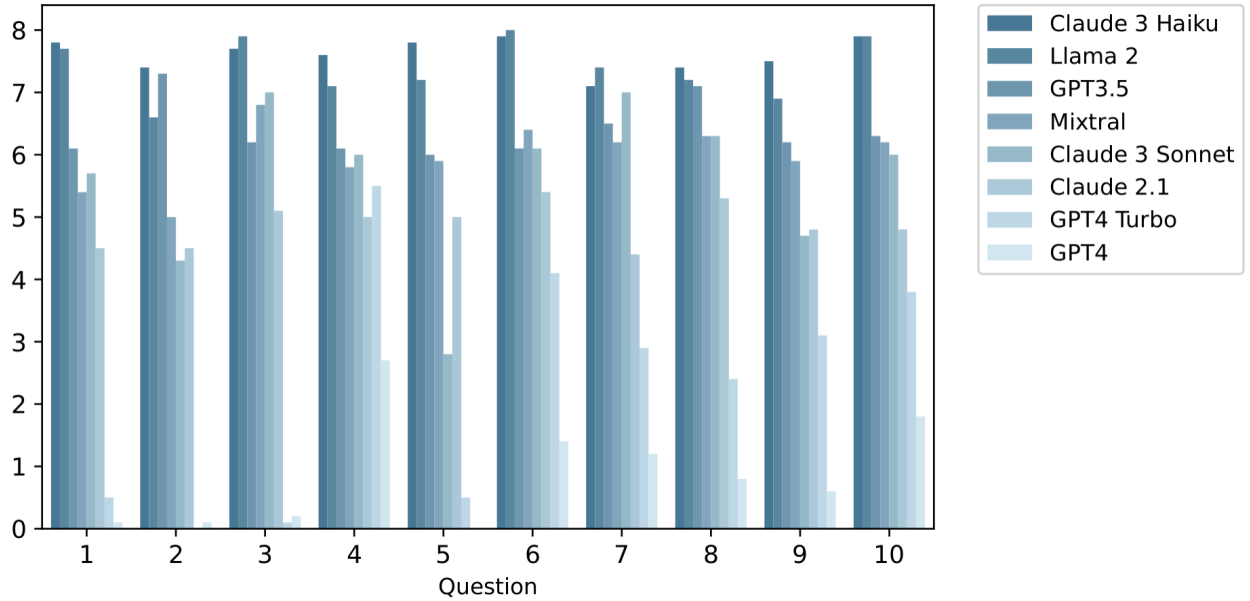
LLMs	Cronbach’s alpha
5	0.707
8	0.629

Note. The Cronbach’s alphas shown here are for the measure used (i.e., after averaging assessments by LLM). For 8 LLMs, all are included. For 5 LLMs, the following are excluded: Claude 2.1, Claude 3 Haiku, GPT3.5.

Table A8. Scores of question specificity (Study 1)

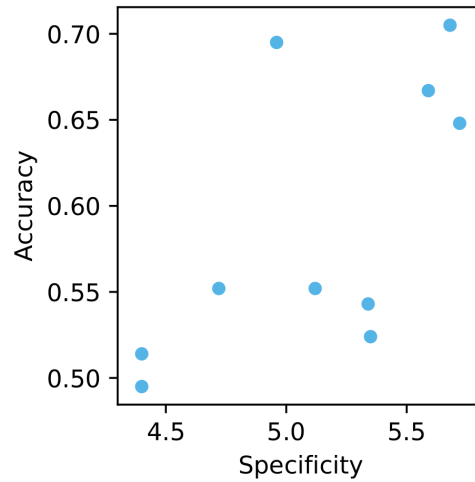
Question	5 LLMs	8 LLMs
1	3.88 (3.39)	4.72 (2.95)
2	3.20 (2.99)	4.40 (2.94)
3	4.40 (3.90)	5.12 (3.19)
4	5.42 (1.64)	5.72 (1.48)
5	3.28 (3.20)	4.40 (2.97)
6	5.20 (2.54)	5.68 (2.14)
7	4.94 (2.74)	5.34 (2.27)
8	4.60 (2.82)	5.35 (2.45)
9	4.24 (2.48)	4.96 (2.24)
10	5.14 (2.37)	5.59 (2.07)

Note. The means are provided with the standard deviation in parentheses.

Figure A2. Average scores of question specificity by LLM (Study 1)

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

Figure A3. Question specificity (from 8 LLMs) and accuracy by question (Study 1)



Note. The correlation is .646. The specificity score is from 8 LLMs.

A.3. Employees

We analyzed the accuracy of perceived origin and also perceived helpfulness by subgroup of employees.

A.3.1. Results

Table A9. 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 A10. Regressions of perceived helpfulness by subgroup (Study 1)

	(1)	
	Perceived helpfulness	
	<i>b</i>	<i>se</i>
Perceived AI	−0.673	0.063***
True AI	0.026	0.043
Senior	0.263	0.289
Perceived AI * Senior	0.008	0.144
Tenure (≥ 3 years)	0.078	0.420
Perceived AI * Tenure (≥ 3 years)	−0.239	0.093*
Questions	FE	
Employees	FE	
<i>n</i> observations	990	
<i>n</i> employees	99	
<i>F</i>	281.1***	
<i>R</i> ²	0.973	

Note. ***: $p < .001$, **: $p < .01$, *: $p < .05$. FE is fixed effects. Model 1 is an OLS regression with robust standard errors clustered by employee. 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: 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.

AI: 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 inferenced. 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.

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	−0.147	0.095	−0.174	0.091
True AI	0.456	0.095***	0.319	0.091***
Labeled AI * True AI	−0.136	0.135	−0.060	0.129
Company: Hershey	3.406	0.095***	2.872	0.097***
Company: General Motors	2.647	0.096***	2.256	0.097***
Company: Nvidia	3.210	0.095***	3.007	0.097***
Company: Amazon	3.407	0.095***	3.047	0.097***
Participants	RE		RE	
<i>n</i> observations	872		872	
<i>n</i> participants	218		218	
<i>LL</i>	−1319.1		−1324.2	

Note. ***: $p < .001$, **: $p < .01$, *: $p < .05$. RE is random effects. Models 1 and 2 are OLS regressions.

References

- Choudhury P, Wang D, Carlson NA, 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, Xia F, .., Zhou D (2022) Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35, 24824-24837.