Discussion Paper Series

IZA DP No. 18442

March 2026

Does AI Cheapen Talk? Theory and Evidence From Global Entrepreneurship and Hiring

Bo Cowgill
Columbia Business School and IZA@LISER

Pablo Hernández-Lagos
Sy Syms School of Business at Yeshiva University

Nataliya Langburd Wright
Columbia Business School

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18442 March 2026

Does AI Cheapen Talk? Theory and Evidence From Global Entrepreneurship and Hiring*

Abstract

Screening human capital based on signals such as job applications or entrepreneurial pitches is crucial for organizations. Signals are often informative insofar as they require differential knowledge and effort to produce. Generative AI (GAI) complicates screening by lowering the cost of producing impressive signals. We model the informational effects of GAI, showing that applicants' access to GAI can increase — but also decrease — an evaluator's screening mistakes. This result depends on how GAI affects experts' signals compared to non-experts'. Using experiments in hiring and startup investing, we estimate that senders' access to GAI (ChatGPT) lowers screening accuracy by 4-9% for employers and startup investors. Consistent with our model, senders' access to GAI also improves screening accuracy in some settings — in our case, among senders from non-English-speaking countries. These results show that GAI can profoundly shape screening accuracy

JEL classification

D82, M51, L26, D83, O33, M13

Keywords

screening, Artificial Intelligence, entrepreneurship, human capital

Corresponding author

Bo Cowgill

bo@bocowgill.org

* All authors contributed equally to this paper. We thank Hemant Bhargava, Nandil Bhatia, Avi Goldfarb, Nan Jia, Ginger Jin, Mike Luca, Neil Thompson, and Emma Wiles for helpful discussion and feedback, as well as seminar participants at Boston University (Digital Initiative), Johns Hopkins (Carey), UCLA (Anderson), and Yeshiva (Sy Syms). We also thank participants at the Columbia Management, Analytics, and Data Conference, the Conference on Field Experiments in Strategy, and the Wharton AI and the Future of Work Conference for comments and suggestions. We are grateful for the Columbia Business School Digital Future Initiative Grant for helping fund this project. Cowgill thanks the Kauffman Foundation Emerging Scholars Program, the Columbia Center for Political Economy, the NET Institute, and the Stellar Development Foundation. Hernandez-Lagos thanks the YU Sy Syms Dean's Research Fund. We thank Noa Rosinplotz and Daniel Creech for excellent research assistance. All errors remain our own.

1 Introduction

Screening talent and ideas is essential for organizations around the world to hire the right employees (Black et al., 2024; Wiles et al., 2023b) or to invest in promising ideas (Bapna, 2019; Scott et al., 2020). Candidates generate various signals that organizations must evaluate to identify the candidate's talent.[1] Such signals are credible insofar as they require differential knowledge and effort to produce. Precisely because of the differential costs of acquiring knowledge and exerting effort, these signals can be informative about the candidate's underlying quality (Spence, 1973).

Generative AI (GAI) presents a challenge to screening: It dramatically reduces the cost of generating signals that evaluators use to assess human capital.[2] Some commentators forecast an "epistemic apocalypse," driven by cheap GAI "deepfakes" in domains spanning politics, hiring, investment, and the broader economy (Habgood-Coote, 2023).

How does GAI impact the accuracy with which evaluators identify a candidate's true level of expertise? In this paper, we develop a method to measure the reduction in evaluators' ability to distinguish expertise from lack thereof. In other words, we quantify *information loss* from GAI. We present a simple model of how GAI affects evaluators' learning and use it to propose a strategy for estimating information loss in the data. The model shows that the informational effects of GAI are ambiguous, contrary to the "epistemic apocalypse" prediction. We derive theoretical conditions in which GAI could decrease the accuracy of screening as well as increase it.

We then measure the size and direction of these effects with experiments in hiring and entrepreneurship. In the experiments, job candidates and entrepreneurs write "pitches,"

[1] Signals are observable pieces of information, such as pitches or cover letters, that enable organizations to infer the expertise of candidates.

[2] In other knowledge tasks, GAI has similarly lowered costs, e.g., Brynjolfsson et al. (2025); Choi and Schwarcz (2023); Eloundou et al. (2024); Felten et al. (2023); Noy and Zhang (2023).

with and without accessing GAI, to convince employers and investors, respectively. We observe each subject's prior experience and expertise, so that we can compare evaluators' assessments of human capital versus the known facts.

On average, candidates' access to GAI causes information losses equivalent to a 4-9% increase in screening errors (the difference between predicted and actual expertise). When candidates can use GAI in our experiments, they produce higher quality pitches, but their signals become more "compressed" and homogenized, making it harder for evaluators to distinguish underlying expertise. We find that when candidates can use GAI for their pitch text, evaluators increase their demand for other signals to evaluate candidates (e.g., background investigations or market research) that are costlier than text by about 6-9%. Demand for these costlier signals reflects lower trust in the pitch as a signal of expertise.

We show that the effect of GAI on screening accuracy depends on a key theoretical parameter: how correlated each candidate's signal boost from GAI and the candidate's expertise are. Over the whole population, we find the covariance between the candidate's signal boost and expertise to be slightly negative (-0.012, $p = 0.072$), implying that GAI helps low-quality candidates in our setting slightly more than high-quality ones, on average. In this case, GAI compresses signals between high and low applicants, thereby pooling them and making it more difficult for evaluators to separate them, which leads to information loss.

However, in different settings, GAI could help high-quality applicants boost their signals, while providing little help to lower-talent candidates. Rather than destroying information, we show that GAI in these settings enhances the signal's informational content, helping evaluators separate experts and non-experts more effectively than without GAI. Although this finding may appear to be a theoretical curiosity, we find empirical evidence of this, particularly when senders are from non-English-speaking countries. For these subjects, GAI helps experts significantly more than non-experts. This makes it easier for evaluators to detect the experts. Screening accuracy is thus higher when certain candidates can use GAI.

Our paper contributes to three bodies of literature. The first is about firms' screening and selection processes (Bapna, 2019; Chan and Wang, 2018; Hui et al., 2023; Kokkodis and Ransbotham, 2023; Li et al., 2020; Wright et al., 2023b). Several recent papers examine how evaluators' adoption of AI and machine learning for screening, both for hiring and investing, impacts screening outcomes (Bhatia and Dushnitsky, 2023; Horton, 2017). Machine-learning algorithms can reduce the cost of evaluating disparate signals while improving screening accuracy (Cowgill, 2020; Li et al., 2020). By contrast, there is less work on how the use of AI by the candidates themselves impacts screening accuracy. The work that does exist suggests both gains and losses to screening errors (Ghose and Ipeirotis, 2010; Hong et al., 2021; Weiss et al., 2022; Wiles et al., 2023b).[3] Our paper helps explain these apparently disparate findings by modeling and measuring the parameter—the correlation between signal gain from AI and expertise—that determines whether GAI increases or decreases such errors.

We also contribute to the literature on the productivity effects of AI (Agrawal et al., 2021; Allen and Choudhury, 2022; Dell'Acqua et al., 2023a; El Sawy et al., 2021; Gans, 2024b; Kim et al., 2024; Jia et al., 2024). Several recent papers suggest that AI boosts productivity, typically for the firm adopting the technology (Brynjolfsson et al., 2025; Choi and Schwarcz, 2023; Eloundou et al., 2024; Felten et al., 2023; Noy and Zhang, 2023; Svanberg et al., 2024).[4] Due to the importance of information transmission, our paper also raises the possibility of productivity *losses* arising from increased screening errors by firms. Our model helps explain findings where GAI raises workers' productivity by supplying them

[3]See also Gans (2024a), Marinescu and Wolthoff (2020), Martin-Lacroux and Lacroux (2017), Sajjadiani et al. (2019), and Sterkens et al. (2023) in labor markets.

[4]The literature is mostly empirical and often relies on the labor augmentation versus substitution framing (for example, Noy and Zhang 2023). One exception is Agrawal et al. (2021), which presents a model where AI helps decision-making by providing information about an external state of nature. The paper analyzes the interplay between information gains from using AI and information transmission between functional units within the firm.

with better information tools, but lowers overall productivity because the people relying on that information trust it less (as in, for example, Kim et al., 2024).

Lastly, this paper sheds light on the context-specific nature of AI. Recent work shows that AI can have different effects across contexts. Some papers find that AI is more helpful for firms or workers at the bottom of the performance distribution (Brynjolfsson et al., 2025; Dell'Acqua et al., 2023b), while other work suggests GAI helps workers at the top more (e.g., Choudhury et al. (2020); Conti and Messinese (2024); Otis et al. (2023)). Our work provides an explanation for why we should expect variation in the distributional effects, for example, across industries or geographies. Our key parameter, the correlation between expertise and GAI signal boost, captures whether workers at the top or bottom of the skill distribution benefit more from the technology than others, and it varies across populations. This suggests that the distributional effects are not inherent to the technology. They depend on the context. Although GAI may seem like a globally-standardized technology, our results suggest that companies may need to pursue context-specific strategies, for instance, when hiring or investing across borders (Bloom et al., 2012a; Carlson, 2023; Gupta and Khanna, 2019; Wright et al., 2023a; Wright, 2025).

In the remainder of this paper, we provide a model of GAI and screening in Section 2. Section 3 discusses the experimental design and data to test the model. Section 4 provides results, and Section 5 discusses and concludes.

2 Theoretical Framework

The model features two players. One is a "sender" who spends effort generating a pitch (such as a cover letter or business plan). Senders with genuine expertise find it easier to generate a high-quality pitch. However, when GAI is available, senders of all types can use generative AI to enhance their pitch. The other player is a "receiver" who interprets the pitch and rewards the sender based on their beliefs about the sender's "true" level of expertise.

As in many real-world settings, the sender has incentives to manipulate the receiver's beliefs, which means that the receiver faces a signal extraction problem. They derive correct beliefs about expertise using an imperfect signal, which contains a mixture of information, noise, and "spam" (akin to, e.g., Holmström, 1999). In the model below, we derive how the signal affects the receiver's beliefs about the sender's expertise. We can then evaluate the accuracy of those beliefs and see how this accuracy changes when senders can use GAI. After the preliminaries below, we summarize notation in Table 1 and proceed to the theoretical results.

2.1 Setup

The Sender. The sender in our framework generates a signal $s \in \mathbb{R}$ by selecting a level of effort $a \in \mathbb{R}$. The signal represents a "pitch" such as an outreach email or business plan. Higher signals require more effort to generate. Effort is costly and convex; exerting a effort reduces utility by $a^2/2$. Senders' signals can also increase or decrease according to the following privately known types:

1. Senders differ in expertise, denoted by $e_n \in \mathbb{R}$. Senders with more expertise find it easier to produce higher signals.
2. Senders vary in how much GAI improves their pitch (when available). We denote this GAI boost as $e_g \in \mathbb{R}$.

The sender privately observes these types (e_n, e_g). We assume they are drawn from a bivariate normal CDF F with strictly positive means $(\bar{e}_n, \bar{e}_g)$, variances (σ_n^2, σ_g^2), and a covariance term σ_{ng}.[5] The joint distribution of e_n and e_g is known to all players. The

[5]The normality assumption on types is for tractability. The main qualitative result in the model holds for a class of discrete distributions as well. See Appendix A4.

covariance between expertise and GAI boost—the σ_{ng} term—can be positive or negative. This term will become a key parameter that drives our qualitative results.

The Signal. A sender of type (e_n, e_g) using effort a can produce a signal:

$$s = a + e_n + G \cdot e_g + \epsilon, \tag{1}$$

where $G = \{0, 1\}$ is a parameter that indicates whether the GAI was available to generate the signal. All terms in Equation 1 are privately known, except G (publicly known). The sender knows her own effort level and private types (e_n, e_g). The variable $\epsilon \in \mathbb{R}$ represents noise and is drawn independently from $\mathcal{N}(0, \sigma_\epsilon^2)$. The noise parameter is known to the sender and is unobserved by all other agents.[6]

The availability of GAI ($G = 1$ or $G = 0$) is an exogenous, publicly known parameter. As motivation for this setup, the state $G = 0$ could represent a pre-GAI era, where the absence of GAI availability is common knowledge. The $G = 1$ could represent a post-GAI era, where GAI is widely known to be available and integrated into content-creation tools. G is publicly known because all parties know which era they are in.[7] Although we raise this pre/post distinction as one motivating interpretation, our model only relies on the assumption that the availability of GAI ($G = 1$ or $G = 0$) is exogenous and publicly known.[8] We relax this assumption in our extensions, allowing the availability of GAI to be unobserved (Appendix A2), and we endogenize GAI use in Appendix A3.

Because of the assumption above, receivers in our setup are *aware* that all senders have *access* to GAI when $G = 1$ (e.g., in our pre/post interpretation, they know that MS Word

[6]We assume that the noise realization is the same for a given sender, irrespective of the availability of GAI.

[7]This characterization may not be a good description of recent years, when GAI remains relatively new. In recent years, some senders might not use GAI despite its availability. Similarly, some receivers may be unaware that senders are using GAI. The use of GAI in today's environment could include considerations we do not model here.

[8]For this reason, we do not mention the pre- and post-GAI distinction to subjects in our experiment.

contains LLM assistance). However, evaluators remain uninformed about any specific candidate's e_g, representing how much GAI helped that specific sender.

Sender's Utility. The sender gains a payment P from the receiver (e.g., being hired or funded) based on the receiver's beliefs about the sender's expertise. The sender chooses a to maximize

$$u_s = P - \frac{a^2}{2}. \tag{2}$$

Choosing to provide higher effort a raises the sender's signal (Equation 1), but there are convex costs for higher effort (Equation 2). Next, we show how the receiver formulates the receiver's payment P.

Receiver. The receiver's *ex-post* value of the sender is equal to their expertise e_n. Qualitatively, these payoffs correspond to a receiver seeking "true expertise" (e_n) and not the ability to artificially enhance a cover letter using GAI (e_g) or through effort (a). The receiver therefore makes inferences from the signal about the sender's expertise and bases the payment P on this expectation.

To recruit the sender, the receiver splits her expected profits (with the sender getting a portion $r > 0$). The receiver's payoff is thus equal to $(1 - r) \cdot \mathbb{E}[e_n|s, G]$, where $\mathbb{E}[e_n|s, G]$ represents the receiver's beliefs about the sender's expertise (given the sender's signal and availability of GAI). The payment to the sender from the receiver is:

$$P = P(s, G) = r \cdot \mathbb{E}[e_n|s, G].$$

The receiver's preferences are common knowledge. Given this setup, the sender has incentives to manipulate the receiver's beliefs about their level of expertise. Table 1 summarizes the notation of the model.

Table 1: **Glossary of Notation**

Symbol	Meaning / Definition	Support / Type
Players, types, and primitives		
e_n	Sender's true expertise (receiver ultimately values this).	$\mathbb{R}$
e_g	Sender-specific GAI "boost" to the signal when GAI is available.	$\mathbb{R}$
F	Joint distribution of (e_n, e_g) (bivariate normal) with means $(\bar{e}_n, \bar{e}_g)$, variances (σ_n^2, σ_g^2), covariance σ_{ng}.	distribution
$\bar{e}_n,\ \bar{e}_g$	Means of expertise and GAI-boost types.	$\mathbb{R}$
$\sigma_n^2,\ \sigma_g^2$	Variances of e_n and e_g.	$\mathbb{R}_{\geq 0}$
σ_{ng}	Covariance between expertise and GAI boost (key parameter).	$\mathbb{R}$
Actions, technology, and noise		
a	Sender's signaling effort; cost $a^2/2$.	$\mathbb{R}$
s	Signal (pitch quality).	$\mathbb{R}$
$s = a + e_n + Ge_g + \epsilon$	Signal technology (Eq. 1).	equation
G	Indicator for GAI availability.	$\{0,1\}$
ϵ	Noise in the signal; $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$.	$\mathbb{R}$
σ_ϵ^2	Variance of noise ϵ.	$\mathbb{R}_{\geq 0}$
E[e$_n$|s, G]	Receiver's posterior mean belief about expertise.	$\mathbb{R}$
P(s,G) = r $\cdot \mathbb{E}[e_n\|s, G]$	Payment to sender.	$\mathbb{R}$
r	Sender's revenue share.	$\mathbb{R}_{>0}$

Notes. This table summarizes notation in the setup of our model. Table A1 in the Appendix expands this table for use in our proofs.

2.2 Analysis

Before sharing our main results in Section 2.3, we now present the key analytical steps. All proofs are in Appendix A. At the core of our results are the receiver's posterior beliefs about the sender's expertise (based on the signal s and availability of GAI G). This term is denoted $\mathbb{E}[e_n|s, G]$. Appendix Lemma A1 derives the formula for this term.

As is well-known for Bayes-updating normal random variables, the posterior belief is a weighted function of the prior and the signal. Our Lemma A1 reflects this, and we use the term $\pi(G)$ to represent the weight placed on the signal. Lemma A1 allows us to derive two important implications about this weight.

Corollary 1 (GAI Changes Weights)**.** *The weight $\pi(G)$ receivers place on the signal can either* *increase or decrease when senders can use GAI.*

Appendix A1 presents results that establish exact conditions when the weight $\pi(G)$ increases or decreases. The weight the receiver places on the sender's signal (i.e., pitch quality) is important because it influences screening accuracy and the other key outcome variables we are interested in.

Corollary 2 (GAI Changes Effort)**.** *The optimal signaling effort a^* can either increase or decrease when senders can use GAI.*

This result suggests that the substitution effect from GAI is theoretically ambiguous. GAI allows higher signals for less effort. This could make senders want to supply more effort to get ahead of other senders. However, senders might also think: If I can get higher signals with less effort, why work harder? This latter possibility is consistent with work showing that AI can lower incentives for effort (Athey et al., 2020; Dell'Acqua, 2022, "falling asleep at the wheel").[9] In an infamous federal lawsuit, for example, an attorney used ChatGPT to enhance his brief (while supplying lower editing effort). The judge discovered fake quotes in the brief and sanctioned the attorney (Weiser, 2023).

Accuracy Loss. The key question for any evaluator is: How far are the receiver's posterior beliefs from the *actual* expertise of a given sender? To answer this question, we examine the squared distance between the sender's true expertise x and the receiver's beliefs about it. Let $f_{\mathcal{B}}(b; x, G)$ denote a PDF representing the receivers' beliefs b about candidates whose true expertise is x. Accuracy loss is defined as the average squared deviations from true expertise:

$$L(x, G) = \int (x - b)^2 f_B(b; x, G) db.$$

[9]The exact conditions for effort to increase or decrease are in the Appendix A1. The conditions coincide with those that make the weight on the signal increase or decrease in Corollary 1 because the optimal effort in equilibrium is proportional to that weight.

In Lemma A2, we derive an explicit and relatively simple formula for the accuracy loss $L(x, G)$ as a function of true expertise x, generative AI availability G, and the distributional parameters. Errors are minimized when a sender's true expertise equals the average expertise in the population: $x = \bar{e}_n$. We also show that the loss averaged across all senders takes a convenient, simplified form below.

Lemma 1 (Average Loss)**.** *Let $\rho(G)$ be the correlation between expertise and the signal when GAI availability is $G = 0, 1$. The receiver's loss averaged over all levels of expertise x, $\mathbb{E}_x[L(x, G)]$, equals the* expected posterior variance of expertise*, i.e. the Bayes risk, or $(1 - \rho(G)^2)\sigma_n^2$.*

Lemma 1 is important because it shows the useful equivalency that the average prediction error equals the variance of the receiver's belief about the sender's expertise (conditional on the signal and GAI availability). Our empirical section will use this equivalence. We will collect data about the receiver's squared-error loss (inaccuracy), as well as the variance (uncertainty) of their beliefs, using both as measures of information loss.

2.3 Main Results

We now proceed to derive our main results. We start by studying accuracy loss: When does GAI help or harm screening accuracy? GAI helps screening if and only if an expert's signal boost from using GAI is typically extremely higher or extremely lower than a non-expert's GAI boost—that is, when the covariance between expertise and GAI boost (σ_{ng}) is extreme.

Proposition 1 (GAI Increases Accuracy)**.** *GAI increases receivers' average accuracy if and only if the covariance σ_{ng} falls outside of the interval (σ^-, σ^+) for constants $\sigma^- < 0$ and $\sigma^+ > 0$. Accuracy strictly decreases otherwise.*

Proposition 1 raises the idea that GAI could *improve* the accuracy of screening. It shows that if the covariance σ_{ng} is extremely high (or low), GAI increases screening accuracy.

Appendix A derives the exact cutoffs for "extremely high" and "extremely low" (i.e., σ^- and σ^+). The intuition is simple. Suppose the covariance between expertise and GAI boost is extremely high. In that case, receivers can easily distinguish experts because experts send very high signals and non-experts very low ones, and so accuracy improves. However, if the covariance is extremely negative, receivers can *still* distinguish experts (i.e., experts are the senders with very low signals), and so accuracy can also improve in this setting.

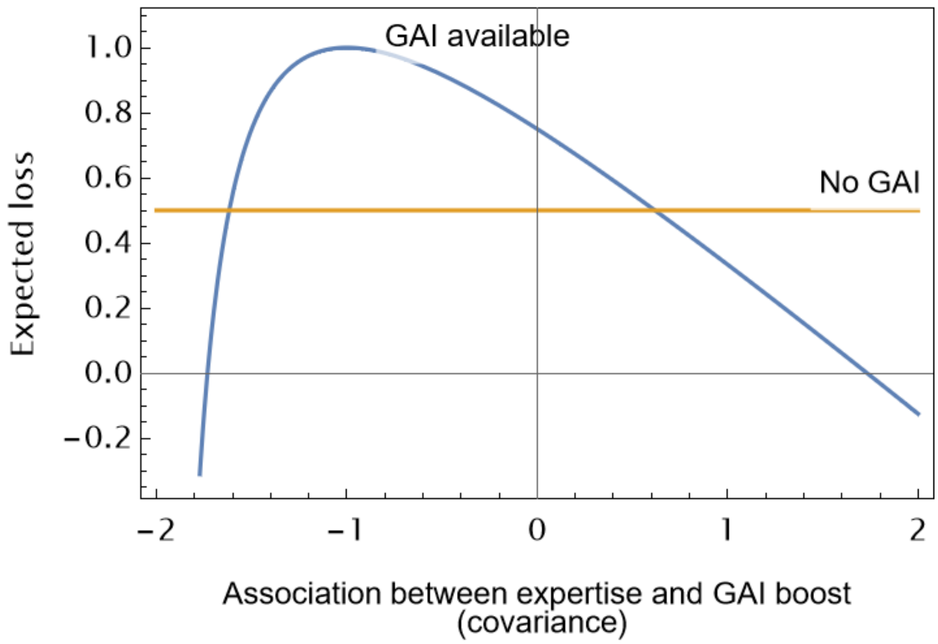
In other words, for GAI availability to help screening, GAI must boost experts' signals significantly more than non-experts' signals, or GAI must hurt experts' signals significantly more than non-experts' signals. In either case, receivers can infer expertise more accurately. By contrast, GAI reduces average accuracy when it benefits experts and non-experts more similarly, particularly when it makes a pitch by a non-expert resemble a pitch by an expert. In such a case, GAI does not help separate expertise. Figure 1 illustrates how GAI makes screening more accurate than without GAI for extreme covariances for some sample parameter values.

In some environments, the opposite of an "epistemic apocalypse" is possible: better learning, not worse. We now discuss whether these "extreme" covariances are realistic (and thus more than a theoretical possibility). In practice, we suspect that the "extremely low" covariance is less likely.[10] However, extremely high (positive) covariances may be possible.

There is evidence for this high positive covariance possibility in a 2023 experiment in AI art by popular science communicator Cleo Abram (Abram, 2022a). Abram hired a visual artist (conventionally trained, pre-GAI) to develop art both without GAI help and later with the assistance of GAI, DALL·E in that case. Simultaneously, Abram herself (no art background) completed the same task, developing art both with and without GAI.

[10] To get an extremely negative covariance, GAI would need to reduce experts' signals (or leave them unaffected entirely), while giving a very high boost to non-experts.

Figure 1: Information Loss Under GAI



Notes: This figure plots the expected squared error loss under GAI (derived in Lemma A2) as a function of different covariance parameters σ_{ng}. GAI makes receivers' inferences more accurate when the covariance between expertise and GAI is extreme. To visualize a specific relationship in this plot, we set variances equal to one, except the variance of GAI boost, which is set to two.

The program then had 3,000 subjects blindly evaluate the professional's and novice's artwork, both with and without GAI. Abram summarized the result, saying "[The GAI] boosted both of us," but

> *"**It didn't feel like the AI was leveling the playing field** between us. It gave me new skills, but **gave [the artist] superpowers**."* – Abram (2022b, emphasis ours).

The conventionally-trained artist developed sophisticated prompts about specific techniques and comparisons from art history, producing much more highly-rated art. The novice's art improved, but she was not able to leverage the power of GAI (DALL·E) without subject-matter expertise. The Abram experiment thus provides an example of GAI helping an expert more than a novice.

The DALL·E experiment was an informal demonstration in a popular science program.

Our experiments in this paper scale the Abram experiment under a controlled setting, using the context of hiring and entrepreneurship rather than AI art.

Our final set of results concerns the pitch. Does the availability of GAI necessarily produce higher signals (i.e., higher-quality pitches)? Does GAI compress the level of signals, homogenizing them, or does it spread them out? We now show that the answer to both questions depends on the covariance σ_{ng}. Let us denote s^* the level of the signal s in equilibrium.

Proposition 2 (GAI Increases Signal)**.** *GAI increases the average level of the signal in equilibrium s^* if and only if $\sigma_{ng} \geq \sigma^S$, where σ^S is a negative constant (and $\bar{e}_g \geq r > 0$). Otherwise, GAI strictly decreases the level of the signal.*

Intuitively, Proposition 2 says that if σ_{ng} is sufficiently high (i.e., experts benefit more), GAI makes pitches look "better"—that is, fancier, higher quality, more polished—or as if more effort has been put into developing them. This "better average pitch" result comes about for two reasons. The first is the direct lift from GAI. With GAI available, the sender's message gets an additive boost $\bar{e}_g$ on average. This lift would otherwise be possible only by working harder.

The second factor is the incentive response. GAI not only makes it look as though the sender worked harder. It also changes the actual amount of effort the sender supplies. When GAI is available, the screeners adjust the amount of trust they place in the signal. Because the sender's equilibrium effort is proportional to how much screeners trust the signal, GAI changes the returns to effort.

If GAI disproportionately helps experts ($\sigma_{ng} >> 0$), the signal is *more informative* about true expertise, receivers increase the weight, and senders work a bit harder. If GAI helps non-experts more ($\sigma_{ng} < 0$), the signal is *less informative*, receivers discount it, and senders optimally reduce effort. This reduction in effort can partially or fully offset the first force: the direct lift. Proposition 2 packages these two forces into a single cutoff $\sigma^S < 0$ (derived in Appendix A).

Our final result is about the *dispersion* or *homogeneity* of the signal. This directly tests the idea that GAI will lead to "compressing" signals, such that experts and non-experts produce similar signal levels. The opposite of compressing is for pitch signals to be widely dispersed and varied, so that they can possibly help distinguish human capital. To study this, we examine the signals' population-wide variance.

Proposition 3 (Signal Dispersion)**.** *The variance of the signal (i.e., the variety of pitch levels in the population) increases with GAI if and only if $\sigma_{ng}/\sigma_g^2 \geq -1/2$, and strictly decreases otherwise.*

Proposition 3 speaks to the widespread claim in popular and academic discourse that GAI will homogenize intellectual output. For example, a June 2025 New Yorker article (Chayka, 2025) titled, "A.I. Is Homogenizing Our Thoughts," cites Agarwal et al. (2025), Kosmyna et al. (2025), and other studies that "suggest that tools such as ChatGPT [...] make our writing less original." Empirical studies by Bommasani et al. (2022), Doshi and Hauser (2024), and Toups et al. (2023) have found similar results, and Kleinberg and Raghavan (2021) and Raghavan (2024) have theorized about diversity and homogenization in AI.

In the stylized setup of our model, Proposition 3 shows that GAI does not necessarily lead to homogenization. It can also lead to higher dispersion and differentiation. The result shows that GAI's impact on the *dispersion* of signals hinges on whether it narrows or widens the gap between experts and non-experts. If GAI helps novices more (negative covariance), low-expertise senders are pulled up while high-expertise senders are pulled less, bunching signals toward the middle. Everyone's signals begin to look alike, compressing the range of quality and making it harder for evaluators to separate candidates.

Conversely, if experts reap the bigger gains (positive covariance), then high-expertise senders get the bigger lift, and dispersion grows. Rather than homogenizing, the signals spread out more, making it easier to distinguish between high- and low-quality candidates. Proposition 3 thus shows that GAI can homogenize signals, though not mechanically: It compresses or diversifies them depending on which group benefits more. As before, the result depends on σ_{ng}. A negative covariance between expertise and GAI boost can actually *decrease* signal dispersion.

Theoretical Extensions. Although the model above illustrates our main theoretical points, we can extend the model to several interesting adjacent questions. Some extensions require simple re-interpretations of the parameters. For example, because the GAI signal boost e_g is unobserved by receivers, one could interpret it as uncertainty about whether GAI was used when available or uncertainty about the boost it gives to the signal when used. Both interpretations lead to the same qualitative results.

The Appendix describes more substantial extensions. Appendix A2 studies exogenous but hidden use of GAI. Appendix A3 endogenizes the use of GAI. Appendix A4 studies binary types. These extensions help apply our setup to other theoretical settings, but the broad message about the pivotal role of the covariance term σ_{ng} arises in all of these models. Section 5.2 discusses several other ways to extend our model, along with some limitations and opportunities arising from our results.

3 Empirical Methodology

Our model suggests that GAI can either increase or decrease screening accuracy. The covariance term—capturing the heterogeneity in how GAI affects signals—is central to the receiver's ability to learn from the availability of GAI. The covariance term may differ across populations and settings.

We now turn to measuring the size and direction of the informational effects of GAI.

The negative epistemic effects of GAI are a common topic in public discourse. Our experiments aim to test and quantify the level of information loss. In applied settings—such as screening job applicants or entrepreneurs—does GAI help or harm screening? How big are the effects?

To answer these questions, we designed and executed a set of experiments outlined below.[11] From the experiments, we can produce reduced-form estimates of how GAI changes signals, beliefs, and the accuracy of screening.

Overview of Experiments. Our experiments took place in Spring/Summer 2023. We recruited subjects into the role of "senders," tasked with developing a signal (pitch). A second set of "receiver" subjects evaluated the signal and interpreted the expertise of the senders. We recruited both sets of subjects from Prolific.[12] In the hiring context, job candidates (senders) wrote cover letters to persuade employers (receivers) to hire them. In the entrepreneurship context, entrepreneurs (senders) wrote pitches to investors (receivers) to get funding for their venture idea.

Senders received instructions to craft pitches intended to signal expertise, regardless of their actual expertise. We asked receivers to evaluate the resulting messages and the expertise of the corresponding senders. Specifically, we tasked the receivers to assess the overall quality of the pitch and to discern which pitches were written by experts. Critically, we have direct measures of senders' *actual* expertise in their respective domains. We can, therefore, assess the accuracy of the receivers' beliefs about the senders' expertise—both when senders can use GAI and not.

Each sender in our design developed four pitches: two in a domain where they have expertise and two in a domain where they do not. Thus, we have a balanced number

[11] The pre-registration may be found here.

[12] Using a controlled context like Prolific is crucial to be able to get pitches from subjects on areas where they do and do not have expertise, as further discussed below.

of pitches by senders with and without expertise. Within each domain, each sender first developed a pitch without using ChatGPT. After writing a pitch without ChatGPT assistance, every sender received instructions that they could use ChatGPT to rewrite or improve the pitch. Thus, we have a balanced number of pitches that could use ChatGPT (and not).

We recruited receivers with prior experience in the relevant domain to evaluate the senders' pitches. Each receiver evaluated eight randomly assigned pitches, each with a 50 percent probability of being written with access to ChatGPT. They were informed whether the pitch was written with the assistance of GAI.[13] Because senders produced multiple pitches in different domains, our specifications can use sender fixed effects. Since each receiver evaluated multiple pitches, we can also use receiver fixed effects. The unit of analysis of our data is thus the combination of sender$\times$receiver$\times$domain$\times$ChatGPT.

In the remainder of this section, we describe our two settings (hiring and entrepreneurship), participant recruitment, treatment compliance, and our full dataset. In Section 3.1, we propose our regression specifications. We choose our two contexts below (hiring and entrepreneurship) to show the robustness of our findings across multiple contexts, and we do not have *ex-ante* reasons to believe that they would generate different effects. We later evaluate heterogeneous effects by setting and find no qualitative differences in the impact of ChatGPT. We provide more detailed information about the design and specific instructions in Appendix C.

Setting #1: Hiring. In the hiring context, we recruit candidates with prior work experience in exactly one of two possible expertise domains (industries): a) data science or b) management consulting. Although each sender has prior expertise in only one domain, they must write a pitch for both roles separately. We task job candidates with developing

[13]The experiments were designed following the interpretation of the model where receivers are told if senders used ChatGPT, rather than only if they had access to ChatGPT. We do this to empirically isolate the impact of a sender's ChatGPT use on receivers' beliefs from the effects of uncertainty around ChatGPT use.

a pitch for their candidacy for a job opening as would appear, for example, in a cover letter or email to a recruiter or employee at the target company.

On the receiver side, we recruit subjects with experience in hiring in the two industries we study. We then task these subjects with reviewing candidates' pitches in the industry of their own prior experience. We inform the receivers about which pitches were developed by senders who could use ChatGPT.

We ask each receiver to evaluate whether the author of each pitch worked in the target role previously. Some senders had worked in the target role and others had not. To answer, subjects used probability intervals. This is our measure of perceived expertise ($\mathbb{E}[e_n|s^*, G]$ in our model). We also ask evaluators to rate the quality of the pitch on a 1–5 scale. We interpret this as a measure of the recruiting pitch's quality in equilibrium (s^*).

Finally, we ask receivers whether they would be willing to pay for a background verification of the sender's expertise conducted by a third party, a proxy that captures demand for costlier signals. Demand for a costlier signal reflects lower trust in the pitch as a signal of expertise (low $\pi(G)$ in the model).

Setting #2: Entrepreneurship. Our entrepreneurship setup has a parallel structure to the hiring one. We recruit subjects with experience as an entrepreneur and who have prior experience in only one of two potential sectors: a) retail or b) education. We then ask the entrepreneurs to write a pitch for a new business idea, both in the sector where they have expertise and the sector they do not. They did not have access to ChatGPT when writing these pitches. After writing them, we gave them access to ChatGPT and informed them they could use it to rewrite their pitches.

As with our hiring setting, ChatGPT and other generative AI tools could be helpful when developing VC pitch materials. Pitch materials often constitute a deck of slides. As of 2024, ChatGPT could directly generate PowerPoint and LaTeX slides. Prior to that, it could generate content (such as graphics or text) that could be helpful for slides. Outside of pitch decks, text remains an important component of VC screening (and often the first step in the screening process).[14]

To review these pitches, we identify investors from around the world with experience in the two sectors. We ask them to evaluate the pitches in the domain of their prior experience. We then elicit their belief about whether the author of each pitch is an entrepreneur with experience in the requisite industry (representing $\mathbb{E}[e_n|s^*, G]$ in our model).

We also ask them to assess the quality of each pitch on a 1–5 scale. This allows us to measure the pitch quality, reflecting the signal level (s^* in our model). As with the hiring scenario, we elicit investors' demand for costlier signals—in this case, a market analysis by a third party that would ask a set of potential customers to evaluate the idea—to validate the quality of the entrepreneurs. As before, demand for costlier signals should be inversely related to how much the investor trusts the signal (the weight $\pi(G)$ in our model).

Heterogeneity: English Language Context. At the heart of our model is a covariance term σ_{ng}. This term represents how much experts (versus non-experts) gain a signal boost from GAI. The parameter σ_{ng} does not appear directly in our data. In Appendix B, we develop a strategy to estimate σ_{ng} from the experimental data.

Different populations may have different covariance terms (σ_{ng}), even for the same GAI technology. This would be consistent with prior work showing differential effects of GAI by the English-speaking-background of users (Brynjolfsson et al., 2025). In our empirical

[14] For example, YCombinator has a text-based application that startups have to fill out even to be considered (`https://www.ycombinator.com/apply`). Even when entrepreneurs send a pitch deck, they often first send a text-based pitch via an introductory email directly to the investor or another contact who then introduces them to the investor. In addition, many entrepreneurs are not able to get a face-to-face meeting with a VC where one could present a deck, and thus make pitches over email (with introductory text, possibly with attachments). The topic of written communication in VC and/or accelerator pitches have appeared in other papers, e.g. Wright et al. (2023b).

context, the GAI is a large language model (ChatGPT) developed originally in English (Radford et al., 2019). Effective LLM use can require follow-up questions and detecting "hallucinations" (Buchanan et al., 2024), which requires sufficient command of the language.

The σ_{ng} parameter could therefore vary by the sender's fluency in English. Effectively using LLMs may be harder for non-native speakers, particularly in domains where they lack expertise. A non-native speaker who does have expertise could possibly compensate for the lack of English with subject matter knowledge. The expertise might have created exposure to some English words at least in this domain, and through this mechanism, the non-native speaking expert could benefit from ChatGPT. For example, an entrepreneur from a non-English-speaking country who has spent a career in the education industry, might be familiar with industry terms in English—such as the name of different educational software—even if they struggle with general English fluency. This industry-specific language knowledge can enable them to better iterate with ChatGPT to create an education startup pitch than can someone without this education industry experience.

For these reasons, the covariance could be positive for senders from outside of an English-speaking background (i.e., ChatGPT will help boost the pitch quality of experts more than that of non-experts). By contrast, the covariance could be the reverse (negative) among native speakers. For these speakers, ChatGPT might help improve the pitches of non-experts more than those of experts. Non-experts from an English-speaking background can ask follow-up questions to ChatGPT until their pitch sounds as convincing as that of an expert. To assess this variance, we recruited participants from around the world. Although our task and subject platform (Prolific) requires some basic English fluency, we collected subjects from both places where English is a primary language and places where it is not.

Participant Recruitment. We recruited senders and receivers from Prolific. We filter senders and receivers based on their industry, occupation, and entrepreneurship experience as provided by this platform. Both senders and receivers get about $12/hour for completing the tasks. Note that our theory does not require that the Prolific experts be, say, world-class experts. It only requires that experts be more qualified than non-experts.

To validate our subjects' relative expertise, we surveyed participants about their prior experience in a variety of domains. Appendix D10 contains our results. Our subjects reported more years of education and work experience in domains relevant to their experimental context. For example, receivers in the entrepreneurial context reported more years of investing experience than other experimental subjects and the broader Prolific pool. This suggests that the "experts" we recruited satisfied the goal of being more qualified than the non-experts. Later, we will directly test whether evaluators rate higher experts' pitches in a blind test.

Compliance. A potential concern with this platform is that senders might be using ChatGPT—even if they are in the control condition. Although we instruct participants not to use ChatGPT in the experimental instructions (see Appendix C), some may have done so anyway. To address this concern, we ask participants at the end of the control (non-ChatGPT) questions whether they actually used ChatGPT in the control condition and exclude those who did. This does not eliminate the possibility that some senders used ChatGPT for the non-ChatGPT conditions. However, this form of non-compliance would bias our results towards zero (by equalizing treatment and control). If this were the case, our results could be interpreted as a lower bound of the true magnitude of effects. We also exclude pitches written by other senders who ignored the experiment's instructions.[15]

[15] For example, some disregarded the prompts altogether by writing a pitch about the retail industry when the prompt instructed them to do so about the education and training industry.

Data. Our final data are on a sender-receiver evaluation level. In total, we have 343 unique senders and 801 unique receivers, about half of each in the entrepreneurship and hiring contexts.[16] Each receiver evaluates eight pitches or cover letters, as illustrated in Appendix C. Four receivers bridged the entrepreneurship and hiring domain, so they completed two sets of evaluations. This leaves our final dataset with 6,440 evaluations.[17] Pitches are randomly assigned to receivers within the same context (either entrepreneurship or hiring). Of our total observations, 3,192 come from the recruitment context, and 3,248 come from the entrepreneurship context. We pool observations across entrepreneurship and hiring in our main paper and test for heterogeneous effects in Appendix D. The senders come from a mixture of English and non-English-speaking countries.[18]

3.1 Regression Specifications

Our specifications measure how GAI changes signals, beliefs, and accuracy. Our main specification is:

$$Y_{idj} = \beta_1 ChatGPT_{idj} + \mathcal{C}_i + \mathcal{R}_j + o_{idj} + \eta, \tag{3}$$

where Y_{idj} is an outcome for sender i on domain d, evaluated by receiver j. We list all outcomes used in our empirical analysis in the next section (Section 3.2), along with their definitions and theoretical analogue.

The variable $ChatGPT_{idj} \in \{0,1\}$ reflects whether the sender/pitching subject could use ChatGPT, corresponding to $G \in \{0,1\}$ in our theory model. $\mathcal{C}_i$ reflects sender fixed effects, and $\mathcal{R}_j$ reflects receiver fixed effects. Because each receiver evaluated multiple

[16] We dropped two of the receivers who were also senders and evaluated at least one of their own pitches.

[17] This total number reflects 797 unique receivers who evaluate 8 pitches each and 4 unique receivers who evaluate 16 pitches each (797 x 8 + 4 x 16 = 6,440). These receivers evaluated pitches in one of the entrepreneurship and one of the hiring contexts. We lose two sender-receiver observations due to incomplete evaluations, but we recover them using natural language processing techniques on some of the dependent variables.

[18] English-speaking countries include: Australia, Canada, South Africa, the U.K., and the U.S. Non-English-speaking countries include countries like: France, Mexico, Poland, and South Korea.

senders, we include pitch order fixed effects (o_{idj}).[19] Standard errors are clustered by receiver. The coefficient of interest is β_1, which indicates whether and how ChatGPT shifts the outcome.

As mentioned before, the model shows how the results depend on a latent parameter σ_{ng} in the data-generating process, representing the covariance between the senders' expertise and the signal increase from ChatGPT. Appendix B contains a strategy to estimate σ_{ng} from our data. In our results section, we mention these estimates of σ_{ng} alongside our reduced-form estimates of our main specification (Equation 3).

Contingency Tables and χ^2 Tests. Although we present most of our results as linear regressions, we also present our results as two-dimensional contingency tables (for example, Tables 3 and 4). On one axis, we have ChatGPT availability vs. not, and on the other axis, we present different levels of the outcome. Within each cell, we present counts of observations. Many of our outcomes are in categorical format, and so we can transparently summarize our data this way. For hypothesis testing, we apply χ^2 tests to the contingency tables. χ^2 tests are advantageous because they are nonparametric—they do not assume normality, linearity, or homoscedasticity. They simply test whether variables are independent by comparing observed frequencies to expected frequencies. Consequently, these tests yield robust, nonparametric insights into the categorical association without risking the biases that arise from regression model misspecification.

3.2 Outcomes

Below, we list and motivate all outcome variables. To link our experiments back to theory, we include a variable name next to each outcome. When interpreting our experiments through the lens of our model, we assume that our empirical measures represent

[19] For each domain setting, we allocated the pitches across eight buckets for each ChatGPT condition, so the order fixed effects also account for differences in expert probabilities within each bucket.

equilibrium outcomes.

Several outcome variables are measured as a 1-5 ordinal rating. In these cases, we use the 1-5 rating as the dependent variable. In some cases, we gave receivers five buckets of choices corresponding to specific numbers or intervals as described below.[20] In these cases, we can either use the 1-5 rating as the dependent variable, or we can use the midpoint of the interval.[21] We mostly use this second approach (midpoints of intervals) to aid interpretation. Qualitatively, our results are robust to either approach, as well as to using the tops or bottoms of the intervals (rather than the midpoint).

In addition, the magnitudes of some effects are not easily interpretable (i.e., squared error loss). To ease interpretation, we present results both in levels and logs. The log results can be interpreted as percentage increases (multiplicative). The five categories below summarize our outcome variables.

1. p**: Receivers' beliefs about sender expertise.** We ask receivers to choose a probability bucket representing how likely a sender is to be an expert. This corresponds to $\mathbb{E}[e_n|s^*, G]$ in our theoretical model (the receiver's posterior mean belief about the sender's expertise). The probability buckets are in intervals of 20% (e.g., 0-20%, 20-40%). We use the midpoint of these intervals as the probability, but our results are robust to using the top or bottom of these intervals.
2. $p \cdot (1-p)$**: Accuracy.** We measure the accuracy of evaluations in two ways: The first approach uses the posterior variance (uncertainty) about expertise, following from Corollary 1. Because receivers' beliefs are probabilities p that the sender is an expert, the variance of this belief is $p \cdot (1-p)$.
3. $(e_n - p)^2$**: Accuracy.** In addition, we study accuracy by taking the square difference between the actual expertise of senders e_n and the receiver's probabilistic prediction

[20] For example, in our question about the sender's expertise, we asked receivers to choose from five buckets of probability in 20% intervals: 0%-20% probability that the sender is an expert, 20%-40%, etc.

[21] For example, 10% for a 0-20% probability bucket.

(mentioned above).

Both 2) and 3) are measures of errors. This corresponds to L (loss) in our theoretical model. Thus, the positive coefficients indicate less accuracy. Because squared errors and variances are difficult to interpret, we also present results in logs (for percentage interpretation). We also measure the level of the signal in two ways. First,

4. s^***: Level of the signal.** We ask receivers to rate the pitch directly using a five point scale. In addition,

5. s^{NLP*}**: NLP Level of the signal.** We also obtain a more "objective" measure of the pitch quality using natural language processing (NLP), and specifically, the Flesch-Kinkaid Reading Score (Flesch, 2007). This score measures the sophistication of written language according to the level of education necessary to understand it.[22] For interpretability, we transform this variable so that higher values represent more sophisticated language.[23] Our NLP variable is 100 minus the raw Flesch-Kinkaid Score.

6. $(s^* - \bar{s}^*_G)^2$**: Variance of the signal.** To measure dispersion, we take the mean ratings of the signal by condition. Let $\bar{s}^*_1$ represent the mean signal under ChatGPT and $\bar{s}^*_0$ represent the mean signal with no ChatGPT. We then subtract each rating s^* from the mean within its condition and square it. The average of these values thus has the interpretation of a variance (corresponding to Proposition 3).[24] We do the same for the NLP rating: $(s^{NLP*} - \bar{s}^{NLP*}_G)^2$.

[22]A concern with the readability measure is that it is only capturing the extremes of the actual quality distribution. Reassuringly, the readability scores in our final sample follow a fairly normal distribution, suggesting that they are capturing quality differences in the middle of the distribution in addition to the extremes.

[23]In the Flesch-Kinkaid Score, higher numbers require less education to understand. The maximum score is 100, which corresponds to a 5th grade reading level, and the lowest score is 0 referring to text readable by professional school graduates. To make "higher" scores better, we transformed the raw Flesch-Kinkaid Score to our NLP variable, which is equal to 100 minus the raw Flesch-Kinkaid Score.

[24]i.e., $Var(X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$, where $\bar{x}$ is the mean.

7. WTP**: Demand for costlier signals**: Finally, we measure evaluators' demand for a signal that is costlier than text, for example, a background investigation in the hiring context and market research in the entrepreneurship context. This variable does not correspond to anything in the model, but is meant to measure subjects' demand for better signals. We elicit this directly in the experiments by asking subjects to choose from a range of dollar values ranging from $0 to $100.

We use these variables as the main outcome variables in Equation 3.

4 Results

Table 2 shows the summary statistics. Many of our main results are in this table, although we later show full regressions with controls. When senders can use ChatGPT, screening accuracy is lower by 4-9%. In other words, screeners are 4-9% more likely to judge recruiters' true expertise incorrectly.

The level of the signal (pitch quality) increases by 4% with ChatGPT when measured as the ratings by receivers. Roughly speaking, the text of the pitches appears more polished and would require higher effort and skill for a human to produce. Table 2 shows that ChatGPT lowers the dispersion of the signal by 54% when using the ratings by receivers and 87% when using the NLP score. These are consistent with GAI availability "homogenizing" the quality of pitches across people of different expertise.

Finally, we find that receivers are willing to pay 8% more on average for additional information in the ChatGPT condition than in the control. This suggests that recruiters sense that the availability of ChatGPT has damaged their screening accuracy and are willing to pay for better signals.

On our main outcome variable (accuracy), Table 3 presents a confusion matrix that sum-

Table 2: **Summary Statistics by ChatGPT Availability**

	With ChatGPT	No ChatGPT	Difference
Sender's True Expertise (0/1)	0.49	0.50	-.0011
Receiver's Belief Sender is an Expert (p)	0.45	0.45	.0037
Variance of Receiver Belief, $p \cdot (1-p)$	0.18	0.17	.0064***
Log(Variance of Receiver Belief)	-1.82	-1.86	.043***
Squared Error: [True Expertise - Receiver Belief]^2	0.31	0.30	.003
Log(Squared Error)	-1.90	-1.99	.092**
Level of Signal (Rating 1-5 by Receiver)	2.91	2.83	.084***
Log(Level of Signal, Rating)	0.96	0.92	.044***
Level of Signal (Natural Language Processing, NLP)	69.95	53.04	17***
Log(NLP Rating)	4.23	3.92	.32***
Variance of Rating	1.45	1.58	-.14***
Log(Variance of Rating)	-1.07	-0.53	-.54***
Variance of NLP Rating	134.31	304.95	-171***
Log(Variance of NLP Rating)	3.51	4.38	-.87***
Receiver Willingness to Pay ($)	39.37	37.27	2.1***
Receiver Willingness to Pay (1-5 Rating)	2.47	2.36	.1***
Log(Receiver Willingness to Pay)	3.39	3.31	.079***

Notes: This table contains summary statistics for ChatGPT and non-ChatGPT messages. $N = 6,438$ total sender$\times$receiver pairs. In $3,270(50.78\%)$ observations, the sender had ChatGPT available to develop their pitch. The stars are given by the p-values of the t-test of a difference in means with zero. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

marizes the accuracy of our screeners' forecasts. The null hypothesis in Table 3's χ^2 test is that the distribution of predicted expertise is statistically independent of the rater's forecasts. The low p-value ($p < 0.001$) represents the rejection of this hypothesis. Together with our evidence, this suggests that our screeners were able to identify experts with a statistically detectable degree of accuracy. We will later quantify how this changes as senders gain access to ChatGPT.

In Appendix D1, we briefly explore differences between the entrepreneurship and hiring settings. Although there are differences in the overall level of outcomes, we find either no differences or small differences in the magnitude and direction of the ChatGPT effects. From here, we now go through our main results in detail using our pooled sample and

Table 3: **Confusion Matrix (Full Experiment)**

	Predicted Expertise					
	$[0, 0.2]$	$(0.2, 0.4]$	$(0.4, 0.6]$	$(0.6, 0.8]$	$(0.8, 1]$	Total
Actual Expert	721	597	707	766	394	$3,185$
Actual Non-Expert	926	648	651	688	340	$3,253$
Total	$1,647$	$1,245$	$1,358$	$1,454$	734	$6,438$

Pearson $\chi^2(4) = 37.3575$, $p < 0.001$

Notes: This table shows a confusion matrix for all predictions made by receivers/screeners in the data. The null hypothesis in the χ^2 test is that the distribution of predicted expertise is statistically independent of the rater's forecasts.

the regression in Equation 3.

4.1 Screening Accuracy

How did accuracy change with ChatGPT? Table 4 reports a contingency table summarizing squared errors by ChatGPT availability. The χ^2 test examines the null hypothesis that the distribution of the error is independent of ChatGPT availability. The low p-value rejects this null (non-parametrically). Appendix Table D3 presents similar contingency tables for absolute error, and Table D4 presents raw errors. Tables D5 and D6 separate out the error distributions for ChatGPT and non-ChatGPT senders. Together, these results consistently show that ChatGPT is associated with screening errors.

Table 5 presents our results about the effects of ChatGPT on accuracy, both with no controls and the full set of controls in Equation 3. The outcomes in this table measure screening errors, so positive coefficients indicate greater inaccuracy. We use both the variance of receiver beliefs as well as squared-error loss as measures of accuracy (as Lemma 1 derives their equivalence).

Table 4: **Squared Error Contingency Table**

	$Error^2$ = (Predicted - Actual Expertise)^2					
	.01	.09	.25	.49	.81	Total
No ChatGPT	703	679	643	599	546	3,170
ChatGPT	617	735	715	686	517	3,270
Total	1,320	1,414	1,358	1,285	1,063	6,440

Pearson $\chi^2(4) = 16.7709$, $p = 0.002$

Notes: This is a contingency table expressing raw counts of observations in all categories. We calculate predicted expertise by using the midpoint of all five probability bins [0-.2], [0.2-.4], [.4-.6], [.6-.8] and [.8-1]. The null hypothesis of the χ^2 test is: The distribution of the errors is identical whether ChatGPT is available or not.

Table 5: **ChatGPT Increases Screening Errors**

	(1) Variance of Receiver Belief $p \cdot (1-p)$	(2) Variance of Receiver Belief $p \cdot (1-p)$	(3) Log Variance of Receiver Belief	(4) Log Variance of Receiver Belief	(5) Error^2	(6) Error^2	(7) Log(Error^2)	(8) Log(Error^2)
ChatGPT	.0064***	.0059***	.043***	.041***	.003	.0011	.092**	.077*
	(.0018)	(.0019)	(.012)	(.012)	(.0067)	(.0074)	(.039)	(.041)
FEs	-	All	-	All	-	All	-	All
R^2	.0023	.26	.0024	.26	.000029	.2	.0009	.22
DV Mean	.17	.17	-1.8	-1.8	.31	.31	-1.9	-1.9
Observations	6,438	6,438	6,438	6,438	6,438	6,438	6,438	6,438

Notes: The table shows results from our regression approach in Equation 3 to study how ChatGPT affects screening errors. As the outcome variable, we use measures of screening errors and regress them on ChatGPT. The columns using all fixed effects are outlined in Equation 3 and include sender, receiver, domain, and order fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Across our outcomes, we find that ChatGPT increases errors. We see this for all our results in logs and in levels for the variance of receiver beliefs measure. In regression format, our result on the level of squared errors cannot be detected at conventional levels, but is highly significant at conventional levels in percentages and logs. We also find differences in squared errors in our most assumption-free tests (Table 4). Across the other outcomes, the decreases in accuracy are detectable at usual significance levels, both with and without fixed effects. The magnitudes of screening errors are comparable to the ones in Table 2, at 4 to 9%.

Why does accuracy go down? In Table D7, we show that ChatGPT does not affect beliefs about the overall level of expertise. Specifically, when senders can use ChatGPT, on average, receivers think the senders have a 45% chance of being an expert. When senders *cannot* use ChatGPT, receivers also think that, on average, the sender has a 45% chance of being an expert (the same overall level). Our standard errors rule out differences greater than one percent.

In a sense, the receivers are correct: The true level of expertise does *not* change between ChatGPT and non-ChatGPT conditions. In both conditions, exactly 50% of senders are "experts." Errors increase under ChatGPT because the receivers label expertise incorrectly (within the same overall amount of expertise). Table D7 suggests that lower accuracy does *not* come from receivers naively believing that more senders using ChatGPT are experts. The overall level is the same. However, the screeners find it more difficult to separate experts and non-experts when senders can use ChatGPT.

We can interpret our findings in Table D7 through the well-known "bias-variance decomposition."[25] In this classic result, receivers' average squared error equals the receivers' bias (squared) plus the variance of their beliefs about the sender's expertise. The results of Table D7 show that receivers are slightly biased about the expertise of senders: They think that senders are 45% likely to be experts, rather than 50% (a statistically significant difference). This bias is one source of screening errors.

However, this bias is constant across the two GAI conditions (Table D7). Instead, what drives the increase in screening errors under GAI is the change in variance: Receivers become more uncertain about expertise under GAI. Columns 1-4 of Table 5 measure this variance directly and show that it increases with GAI. Columns 5-8 show the same qualitative result when the variance is measured as the squared difference between the re-

[25] E.g., see Hastie et al. (2017); James and Stein (1961); Stein (1956).

ceiver's beliefs and the sender's true expertise, although the estimates are less precise.

4.2 The Level of the Signal (Pitch)

We now study how ChatGPT affects the level of the signal. In our model, higher level signals require more effort to produce, but less so for experts. Before we study the effect of ChatGPT, we first use our data to assess a key assumption of the model: that expert subjects generate higher quality signals because it is easier for them to do so (e.g., Equation 1). In Table 6, we show that subjects with prior experience indeed produce pitches

Table 6: **Expert Signal Quality is Higher**

	(1) Signal Level (Rating)	(2) Signal Level (Rating)	(3) Signal Level (Rating)
Expertise	.16***	.13***	.18***
	(.031)	(.045)	(.046)
FEs	All	All	All
Senders' sample	All	ChatGPT	No ChatGPT
R^2	.37	.51	.55
DV Mean	2.9	2.9	2.8
Observations	6,438	3,239	3,135

Notes: In this table, we predict the level of the signal using whether senders are experts. Column 1 shows the full sample. Column 2 shows pitches that can use ChatGPT. Column 3 shows pitches that cannot use ChatGPT. Robust standard errors (in parentheses) are clustered at the receiver level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

that receivers rate as higher quality (column 1), both in our ChatGPT sub-sample (column 2) and outside it (column 3).

Although we see a positive relationship between our measure of expertise and the ratings, some readers may ask why the relationship is not stronger. Decades of prior research about selection (starting at least with Dawes, 1979) suggests that detecting talent in hiring and entrepreneurship settings is very hard. In fact, many employers and investors are relatively bad at it (McDaniel et al., 1994; Nanda et al., 2020; Sackett et al., 2022; Schmidt and Hunter, 1998; Scott et al., 2020).

What textual features did evaluators reward with higher signal levels and beliefs of expertise to pitches? To assess this, we conduct topic modeling on the sender pitches in each context using Latent Dirichlet Allocation (LDA). The LDA model detects clusters of words in the corpus of text that constitute different topics (Blei et al., 2003; Pritchard et al., 2000). We then use ChatGPT to help us label these topics based on the cluster of words. This approach allows us to categorize the pitches within each context into categories of topics.

Our results in Appendix D4 validate the receivers' judgments. Receivers give higher ratings to pitches that discuss topics appropriate for the context (education, retail, data science, or management). For example, in the consulting recruitment context, evaluators give higher ratings to cover letters that mention strategic consulting and client management topics and give lower ratings to ones that mention data science topics.

ChatGPT and Signal Levels. We now turn to the effect of ChatGPT. Table 7 shows how ChatGPT impacts the level of the signal, both with controls and without. The results show that ChatGPT increases the level of the signal by 4-6% when using the evaluators' rating as the measure of level (columns 1–4).

When using the NLP-created measure (columns 5–8), we find larger percentage effects of around +32%. The base rate in the non-ChatGPT control group is 53 (a Flesch-Kinkaid Score of 47), which corresponds to the reading level of a high-school senior. The average in the ChatGPT condition is 70 (or 30 on the Flesch-Kinkaid scale), which is around the reading level of a college senior.

In percentage terms, the result is a higher increase (+32%) than that for the receivers' ratings of the signal (+4-6%). While the NLP-created measure captures language sophistication, receivers seem to believe there is more to a good pitch than the sophistication of

Table 7: **ChatGPT Increases the Level of the Signal**

	(1) Rating by Receiver	(2) Rating by Receiver	(3) Log Rating by Receiver	(4) Log Rating by Receiver	(5) NLP Rating	(6) NLP Rating	(7) Log NLP	(8) Log NLP
ChatGPT	.084**	.12***	.044***	.056***	17***	17***	.31***	.32***
	(.034)	(.031)	(.014)	(.014)	(.41)	(.34)	(.0075)	(.0065)
FEs	-	Full	-	Full	-	Full	-	Full
R^2	.0012	.37	.0018	.37	.25	.7	.26	.7
DV Mean	2.9	2.9	.94	.94	62	62	4.1	4.1
Observations	6,438	6,438	6,438	6,438	6,440	6,440	6,440	6,440

Notes: The table shows results from our regression approach in Equation 3 to study how ChatGPT affects the level of the signal. As the outcome variable, we use the level of the signal (as measured by evaluators' ratings and NLP) and regress it on ChatGPT. The columns using all fixed effects are outlined in Equation 3 and include sender, receiver, domain, and order fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

language. When pitching for a job or startup idea, other aspects (e.g., qualifications and ideas) presumably also matter. ChatGPT may not be able to provide these as effectively, and thus the effect of ChatGPT on the receivers' ratings is lower than that on the NLP score.

But how were ChatGPT pitches different? We can use our LDA topic models to study this. Appendix D5 shows that ChatGPT shifts the topics of the pitches. For example, in the education startup context, ChatGPT pitches mention "Specialized Training" significantly less than non-ChatGPT pitches. Our results on this (summarized in Appendix D5) generally suggest that ChatGPT shaped the content of the pitch rather than simply correcting spelling and grammar.

Although the signal was higher for pitches that could use ChatGPT, did receivers believe that the ChatGPT senders had more expertise? Table D7 suggests that they were not "fooled:" Average beliefs about the level of expertise did not increase. Together, these results suggest that the pitches looked better in the ChatGPT condition (Table 7), but they also suggest that receivers discounted the better pitches, attributing the difference to ChatGPT.

4.3 Dispersion of the Signal

ChatGPT increases the level of signals. But how does it affect the dispersion of the signal? Table 8 shows that ChatGPT homogenizes or "compresses" the level of the signal, lowering the variance between different senders' pitches. Using the receivers' rating of signals, the variance decreases by about 9%. In logarithms, it decreases by up to 56%. Using the NLP measure, the variance decreases by up to 92%. Our signal results fit intuitively with our overall result about screening errors, suggesting that screeners use varied signals to distinguish expert from non-expert candidates or entrepreneurs.

Together, our results suggest that ChatGPT raised the average level of the signal while significantly reducing variation in it. Before ChatGPT, signals were low and highly dispersed (i.e., many high signals, many low signals). After ChatGPT, signals are higher but more similar.

Table 8: **ChatGPT Lowers Dispersion of Signals**

	(1) Var. of Rating	(2) Var. of Rating	(3) Log Var. of Rating	(4) Log Var. of Rating	(5) Var. of NLP	(6) Var. of NLP	(7) Log Var. of NLP	(8) Log Var. of NLP
ChatGPT	-.14***	-.13***	-.54***	-.56***	-171***	-186***	-.87***	-.92***
	(.041)	(.043)	(.057)	(.062)	(16)	(16)	(.054)	(.053)
FEs	-	All	-	All	-	All	-	All
R^2	.002	.27	.015	.23	.019	.51	.036	.48
DV Mean	1.5	1.5	-.8	-.8	218	218	3.9	3.9
Observations	6,438	6,438	6,438	6,438	6,440	6,440	6,440	6,440

Notes: The table shows results from our regression approach in Equation 3 to study the dispersion of signals. As the outcome variable, we use the variance of the signal as measured by evaluators' ratings and NLP and regress it on ChatGPT. The columns using all fixed effects are outlined in Equation 3 and include sender, receiver, domain, and order fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.4 Demand for Costlier Signals

Our final result is about how screeners might react to the loss of information. Our results above suggest that screeners trust pitches less when senders can use ChatGPT. For that reason, they may prefer an alternative way to screen altogether. To test this, we ask evaluators how much they would be willing to pay for costlier signals, such as a background investigation (in the recruitment setting) or market research (in the entrepreneurship setting). Table 9 reports the results. We find that receivers increase their willingness to pay by about 6-9% when senders can use ChatGPT regardless of the specification. For example, from columns 1 and 2, this translates into an additional $2.10–$2.40 per pitch. Consistent with the reduced accuracy of the pitches, this result suggests that receivers find the pitches less informative.

Table 9: **ChatGPT Increases Demand for Costlier Signals**

	(1) WTP	(2) WTP	(3) Log WTP	(4) Log WTP	(5) WTP, 1-5 Scale	(6) WTP, 1-5 Scale	(7) Log WTP, 1-5 Scale	(8) Log WTP, 1-5 Scale
ChatGPT	2.1***	2.4***	.079***	.089***	.1***	.12***	.055***	.061***
	(.73)	(.62)	(.022)	(.019)	(.036)	(.031)	(.016)	(.014)
FEs	-	All	-	All	-	All	-	All
R^2	.0016	.47	.0023	.46	.0016	.47	.0022	.47
DV Mean	38	38	3.3	3.3	2.4	2.4	.72	.72
Observations	6,438	6,438	6,438	6,438	6,438	6,438	6,438	6,438

Notes: The table shows results from our regression approach in Equation 3 to study the demand for costlier signals. As the outcome variable, we use receivers' willingness to pay for costlier signals, which we regress on ChatGPT. The columns using all fixed effects are outlined in Equation 3 and include sender, receiver, domain, and order fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

If screeners migrate away from text and towards different signals, what does this mean in context? Appendix D6 reports the results of a supplemental survey sent to our subjects as well as a broader pool of Prolific subjects. We find that in the recruitment context, receivers most commonly prefer interviews. In the entrepreneurship context, receivers most commonly prefer market research. After these signals, receivers also value those related to the candidate's personal background.

4.5 Covariance: How much is Expertise Correlated with ChatGPT Signal Boost?

We have thus far shown that ChatGPT: 1) decreases screening accuracy, 2) increases signal levels, and 3) decreases signal dispersion. According to Propositions 1–3, these results suggest that the covariance σ_{ng} should be negative because of 2) and 3), but its magnitude should not be too large because of 1). Otherwise, GAI would improve non-experts' signals so much that they would be more easily distinguished from experts than without GAI.

Applying our strategy for estimating σ_{ng}, as outlined in Appendix B, reveals a negative covariance—with a point estimate of -0.012 and a bootstrapped standard error of 0.0066. Our p-value for testing this estimate against zero is 0.072. The negative (and significant) value suggests that overall, ChatGPT helped non-experts more than experts, and the GAI-induced boost makes non-experts more similar to experts.

The above estimate of σ_{ng} uses our entire dataset. In our experimental design, we discussed the possibility that σ_{ng} could differ by population, particularly according to English familiarity. Among subjects from non-English speaking backgrounds, we hypothesized that experts could use ChatGPT more effectively (than non-experts) to boost their pitches.

In Appendix D7, we examine the use of ChatGPT among non-English speakers. Although we do not know how much time each subject spent iterating with ChatGPT, we can measure the differences between their pitch before and after being able to use ChatGPT. This suggests different intensity of usage and possibly a higher perceived value in the use of ChatGPT. In Table D18, we look at all subjects from non-English backgrounds. Among these subjects, pitches by experts changed more after being able to use ChatGPT (compared to those by non-experts). This is consistent with our hypothesis that among non-English speakers, experts can more effectively use ChatGPT.

We now estimate the covariance by these sub-populations. We find that for senders from non-English-speaking countries,[26] σ_{ng} was +0.056 (positive), with a bootstrapped standard error of 0.013 and a p-value below 0.001 when tested against zero. This indicates that ChatGPT helps experts more than non-experts and thus could possibly *increase* screening accuracy. By contrast, for senders who *do* come from English-speaking countries, we find a more strongly negative estimate of σ_{ng}: a point estimate of -0.034 with a bootstrapped standard error of 0.0075 and a p-value below 0.001 when tested against zero.

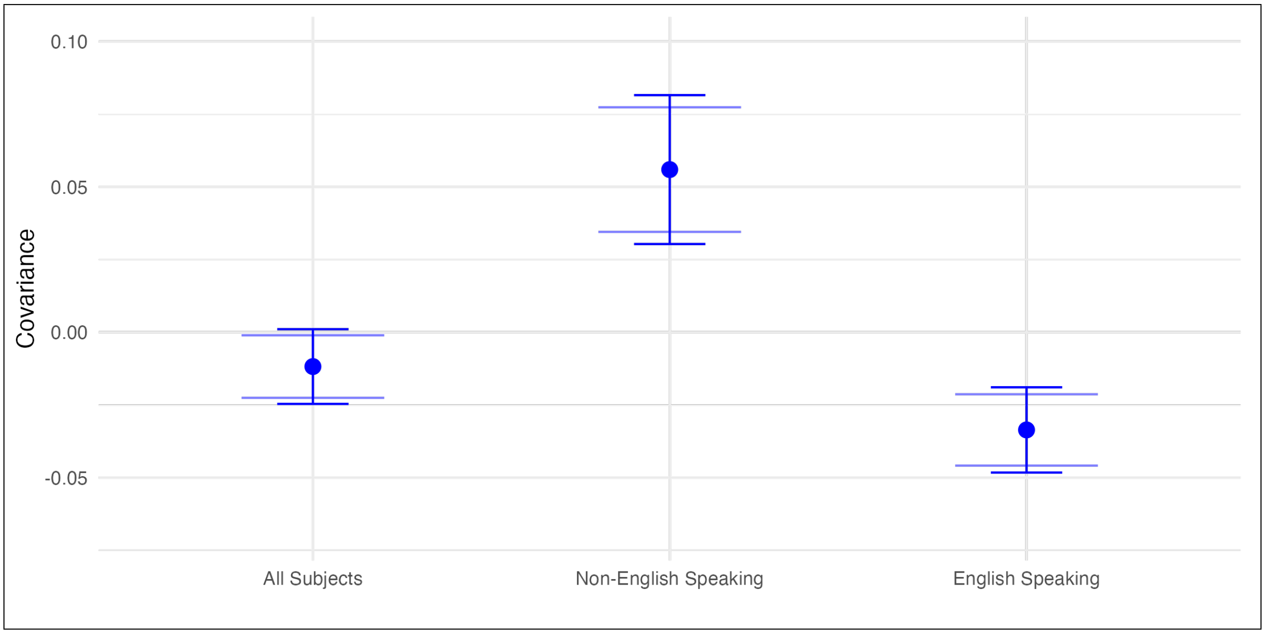
Figure 2 illustrates the differences in our estimate of σ_{ng} across populations. Given that the sign of the covariance differs across these two populations, we now examine heterogeneity in our reduced-form treatment effects of ChatGPT.

4.6 Heterogeneity: ChatGPT and Non-English Sender Contexts

From here, we can now revisit out treatment effects. In Table 10, we study our main outcomes and add an interaction term for non-English-speaking sender countries. We find some evidence that ChatGPT introduces *gains* to accuracy when screening senders from non-English-speaking contexts, for whom experts' signals increase more than non-experts' (σ_{ng} is positive). Consistent with Proposition 1, the coefficient on the interaction is negative and significant at conventional levels in column 1 for our accuracy outcome. The

[26] In robustness tests, whether the sender is from an English-speaking country highly correlates (about 0.8) with English being the sender's first language based on demographic data that Prolific collects about subjects. However, several subjects withheld data about their first language from Prolific, so we use the country—which is available for all senders in our data—as the main measure in our results.

Figure 2: **Covariance Estimates**



Notes: This figure shows 90% and 95% confidence intervals for our covariance estimates across the three populations we study. The covariances are estimated using the econometric strategy detailed in Appendix B.

results are qualitatively similar, though not significant, when using variance of receiver beliefs as a proxy of accuracy (column 2).

However, ChatGPT's impact on the average signal level is lower for senders from non-English-speaking contexts, using the subjective and NLP ratings (columns 3-4), which is inconsistent with Proposition 2. ChatGPT's impact on the signal's dispersion is higher in magnitude but not statistically different in non-English-speaking contexts relative to others, which is qualitatively consistent with Proposition 3 (columns 5-6). Lastly, the impact of ChatGPT on willingness to pay for costlier signals is lower in magnitude, but not statistically significant, in non-English-speaking contexts. This is qualitatively consistent with our theory (column 7).

Together, these analyses suggest that the effects of GAI are not unidirectional. There are cases when GAI can increase screening accuracy. Consistent with the model, when experts benefit substantially more than non-experts—as we see among senders in non-

Table 10: **ChatGPT and Non-English-Speaking Sender Contexts**

	(1) Log(Error^2)	(2) Log Variance of Receiver Belief	(3) Log Rating by Receiver	(4) Log NLP	(5) Log Var. of Rating	(6) Log Var. of NLP	(7) Log WTP
ChatGPT	.17***	.051***	.069***	.34***	-.6***	-.96***	.1***
	(.049)	(.014)	(.015)	(.0075)	(.07)	(.062)	(.022)
ChatGPT× Non-English	-.34***	-.038	-.048*	-.092***	.15	.14	-.048
	(.096)	(.027)	(.028)	(.012)	(.13)	(.12)	(.04)
FEs	All	All	All	All	All	All	All
R^2	.22	.26	.37	.7	.23	.48	.46
DV Mean	-1.9	-1.8	.94	4.1	-.8	3.9	3.3
Observations	6,438	6,438	6,438	6,440	6,438	6,440	6,438

Notes: The table shows results from our regression approach in Equation 3, adding an interaction between Non-English-speaking subjects and ChatGPT. We study all of the main dependent variables. The columns using all fixed effects are outlined in Equation 3 and include sender, receiver, domain, and order fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

English-speaking contexts—ChatGPT helps evaluators screen candidates.

4.7 Robustness Analyses

There are a few ways we can extend our experimental results. For example, our experiments were designed to capture both within-subject and between-subject variation. However, we can modify our analysis to capture only the between-subject variation. We implement this analysis in Appendix D9 and find qualitatively similar results as in our full sample.

In addition, we can study the effects of ChatGPT on rankings. Most of our results so far have studied the level of signals or perceived expertise. However, candidates often compete in a ranking. Appendix D8 studies the effects on rankings. We also find that expertise increases the ranking of the level of the sender's signal by 2.8-3.5 percentiles. Experts also have 4.0-4.8 percentiles higher perceived expertise. However, ChatGPT increases the sender's ranking by only about 2 percentiles and has no effect on perceived expertise. These results are generally in line with our findings in levels (rather than rankings).

5 Discussion and Conclusion

Our paper aims to contribute a deeper conceptual understanding of GAI's impact on information loss and empirical estimates of the degree of information loss, at least in a few key settings. We propose a model of the data-generating process behind GAI information loss, as well as a research design for measuring it. When we implement this design empirically, we can quantify both the degree of information loss (of 4-9%) and the magnitude of the parameters generating that change (a negative $\sigma_{ng} = -0.012$).

5.1 Theoretical Implications

Our study offers a theory, rooted in the covariance parameter, of when GAI leads to information loss. In doing so, it helps reconcile prior work on the impact of GAI on screening that suggests either gains or losses (Gans, 2024a; Wiles et al., 2023a) to screening accuracy when candidates can use GAI. Ultimately, the study tackles a crucial problem that this literature outlines: "the problem of deciding what's real or not" (Horton, 2024; Campante et al., 2025).

By showing how generative AI correlates with expertise in different ways across geographies, the paper also helps explain seemingly conflicting findings on the distributional effects of AI. Indeed, some papers find that AI is more helpful for firms or workers at the bottom of the performance distribution (rather than the top, Brynjolfsson et al., 2025; Dell'Acqua et al., 2023b). Other work reveals the opposite: GAI helps those at the top more (Choudhury et al., 2020; Conti and Messinese, 2024; Otis et al., 2023). Our model embraces this distributional distinction as a parameter that profoundly shapes the accuracy of screening. In doing so, it reveals the context-specific nature of GAI, which makes it important for firms to pursue differentiated strategies across markets.

This work further contributes to research on how technology shapes firm growth and innovation disparities across geographies (Bloom et al., 2012b; Nagaraj et al., 2020; Tambe, 2014). Our work suggests that GAI can potentially reduce the gap between English-speaking hubs that have historically held a competitive advantage (Bloom et al., 2012b; Conti and Guzman, 2023; Kerr and Robert-Nicoud, 2020) and others. Because GAI improves screening accuracy in non-English-speaking contexts, it can enable higher-quality companies from there to raise venture capital and choose expert talent. Access to this funding and talent can be crucial for growth and innovation (Conti and Guzman, 2023).

5.2 Limitations and Opportunities

Our study is not without its limitations, ones that can open the door to several lines of future research.

Empirical Setting. We chose our empirical setting for its ability to support a clean experiment, though it has limitations. For example, our measure of expertise captures a binary rather than continuous distribution, our subjects participated in an artificial intelligence exercise that may not fully capture behavior in higher-stakes situations, and all subjects in the GAI condition had access to the technology. This paper seeks to be a Wave 1 study, focusing on establishing initial causality and producing first tests of theory (List, 2020). Future work may explore this framework with finer-grained measures, in field settings with costlier signals like grant funding applications, and in contexts where AI adoption is unequal (McElheran et al., 2024; Otis et al., 2024).

Vertical vs. Horizontal Differentiation. The model and empirics focus on vertical differentiation only (high versus low-quality candidates, experts versus not). However, many labor markets feature match-specific preferences or productivity. Future research could study how GAI impacts signaling on "horizontal" features where market participants have different tastes. This setting may be promising for GAI, although there may still be some incentive to "fake" one's horizontal type in some cases.

Endogenous σ_{ng}. In our model, the covariance parameter is an exogenous parameter. However, in real life, firms can shape the σ_{ng} of their GAI products. GAI firms can choose to make their technology help low-expertise workers evade screeners, or they can build GAI to help higher quality workers signal their expertise. Deciding which customers to prioritize is a classic entrepreneurial dilemma (Gans et al., 2019). Endogenizing σ_{ng} is a valuable area for future research.

Multidimensional Screening. In our model, the reviewer is screening for one characteristic. We call this "expertise," however, it can generalize to screening for any single variable. Of course, in many settings, reviewers are assessing multiple characteristics (e.g., both technical expertise as well as the ability to speak clearly to the client). In these settings, a single signal could be informative about multiple traits. GAI could boost signals on some of these traits while be neutral on others. Although we do not explicitly show this, it is a natural extension of our model.

Multi-tasking. Investors and employers evaluate candidates on the basis of one dimension of expertise in our model and experiments. However, they might want to screen candidates on two dimensions, one important and a second less important (Holmstrom and Milgrom, 1991). GAI may asymmetrically dilute the signal quality for the important one, leading firms to focus on the second one that is now relatively more measurable, though less crucial for productivity. This would reflect the "The Folly of Rewarding A, While Hoping for B" (Kerr, 1975). Future work may explore this potential implication.

Static vs. Dynamic Model. Our model studies one-shot learning and transfers. In some real world situations, a screener who is disappointed with their employee's on-the-job performance could do something to discipline the candidate. The threat of such ex-post punishment might constrain the incentives to "fake" expertise through ChatGPT.

How GAI Changes Pitches. Our study reveals that GAI improves the average level of the signals. This could be for a variety of reasons. For example, GAI can prompt subjects to simply think more clearly and edit their original pitches. Or GAI can offer new information, for example, about problems in the education sector or skills crucial for a consultant, that addresses subjects' pre-existing knowledge gaps. Precisely disentangling refinement from new knowledge is difficult. Future work can explore the channels behind GAI's boost to pitch quality.

Alternative Signals. Our research additionally suggests that evaluators demand costlier signals than text when ChatGPT is available. But what is the nature of these costlier signals? While our supplementary survey suggests an important role for market research and job interviews, we do not directly observe this choice in our data, nor incentivize it. Will screeners turn to nuanced investigations of individual quality, for example, through a full background investigation or market analysis? Will they leverage more advanced "spam-proofing" techniques like cryptography with existing text-based signals (Horton, 2024)? Or will they focus on demographic variables and protected categories as occurred after US "ban the box" reforms that eliminated criminal history signals (Agan and Starr, 2018; Doleac and Hansen, 2020)? Future work may investigate these possibilities.

Using GAI at Work. Finally, the assumption in our model is that employers want to hire "true" experts, rather than effective users of GAI. There are, however, a variety of ways that GAI could substitute for some (or all) parts of human capital. This may result in employers screening for expertise less and screening instead for other qualities. Thus, the nature of expertise could change with GAI. Understanding these future screening scenarios has important implications for accuracy and bias in hiring, entrepreneurial finance, and many other economic transactions. We hope that future work investigates these possibilities using incentive-based experiments.

5.3 Practical Implications

Our findings have implications for talent evaluators, candidates, and GAI companies. Our average information loss results suggest that as GAI permeates all forms of media (text, pictures, video, audio, etc.), these conventional signals could become useless for screening.[27] Still, GAI could help with screening if it enables high-quality candidates to improve their signals much more than low-quality ones. In any case, our results suggest that evaluators could seek alternative screening approaches. This could create a new market for products that can evaluate human capital and resist GAI's "garbling" of traditional signals.

Despite the "epistemic apocalypse" intuition, we show that GAI does not necessarily destroy information. If GAI disproportionately helps experts—as the DALL·E experiment and our non-English-speaking subjects indicate—then GAI can actually increase the accuracy of screening. While these instances may be relatively rare in the current environment, GAI companies could shape this outcome in the future. These companies have some control over whether their products are more helpful for novices or experts

[27] Online ratings and reputation are other signals that may have lost informational content over time through inflation, particularly when reviews are public and reciprocal (Bolton et al., 2013; Filippas et al., 2018; Nosko and Tadelis, 2015).

(σ_{ng}), and thus may be able to shape the covariance terms and ultimate signaling value of the content that their software produces.

Ultimately, screening extends far beyond the hiring and startup settings that we studied. News, consumers, educators, students, apartment renters, shoppers, singles, educators, medical patients (and more) all heavily rely on screening and signaling. Our paper suggests that GAI could profoundly change how these agents learn and screen. Many signals are useful for evaluators because of their costliness. However, their informational content fundamentally changes if GAI makes these signals cheaper, with implications for a variety of transactions across the global economy.

References

Abram, Cleo, "The REAL Fight Over AI Art," 2022. Available at: https://www.youtube.com/watch?v=NiJeB2NJy1A.

_ , "Tweet by Cleo Abram," September 2022. Accessed: 2024-11-19.

Agan, Amanda and Sonja Starr, "Ban the box, criminal records, and racial discrimination: A field experiment," *The Quarterly Journal of Economics*, 2018, *133* (1), 191–235.

Agarwal, Dhruv, Mor Naaman, and Aditya Vashistha, "AI suggestions homogenize writing toward western styles and diminish cultural nuances," in "Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems" 2025, pp. 1–21.

Agrawal, Ajay K, Joshua S Gans, and Avi Goldfarb, "Ai adoption and system-wide change," Technical Report, National Bureau of Economic Research 2021.

Allen, Ryan and Prithwiraj Choudhury, "Algorithm-augmented work and domain experience: The countervailing forces of ability and aversion," *Organization Science*, 2022.

Athey, Susan C, Kevin A Bryan, and Joshua S. Gans, "The allocation of decision authority to human and artificial intelligence," in "AEA Papers and Proceedings," Vol. 110 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2020, pp. 80–84.

Bapna, Sofia, "Complementarity of signals in early-stage equity investment decisions: Evidence from a randomized field experiment," *Management Science*, 2019.

Bhatia, Abhishek and Gary Dushnitsky, "The Future of Venture Capital? Insights Into Data-Driven VCs," *California Management Review*, 2023.

Black, Ines, Sharique Hasan, and Rembrand Koning, "Hunting for talent: Firm-driven labor market search in the United States," *Strategic Management Journal*, 2024.

Blei, David M, Andrew Y Ng, and Michael I Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, 2003, *3* (Jan), 993–1022.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen, "Americans do IT better: US multinationals and the productivity miracle," *American Economic Review*, 2012.

_ , _ , and John Van Reenen, "Americans do IT better: US multinationals and the productivity miracle," *American Economic Review*, 2012, *102* (1), 167–201.

Bolton, Gary, Ben Greiner, and Axel Ockenfels, "Engineering trust: reciprocity in the production of reputation information," *Management science*, 2013, *59* (2), 265–285.

Bommasani, Rishi, Kathleen A Creel, Ananya Kumar, Dan Jurafsky, and Percy S Liang, "Picking on the same person: Does algorithmic monoculture lead to outcome homogenization?," *Advances in Neural Information Processing Systems*, 2022, *35*, 3663–3678.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond, "Generative AI at work," *The Quarterly Journal of Economics*, 2025, p. qjae044.

Buchanan, Joy, Stephen Hill, and Olga Shapoval, "Chatgpt hallucinates non-existent citations: Evidence from economics," *The American Economist*, 2024.

Campante, Filipe R, Ruben Durante, Felix Hagemeister, and Ananya Sen, "GenAI Misinformation, Trust, and News Consumption: Evidence from a Field Experiment," Technical Report, National Bureau of Economic Research 2025.

Carlson, Natalie A, "Differentiation in microenterprises," *Strategic Management Journal*, 2023, *44* (5), 1141–1167.

Chan, Jason and Jing Wang, "Hiring preferences in online labor markets: Evidence of a female hiring bias," *Management Science*, 2018, *64* (7), 2973–2994.

Chayka, Kyle, "A.I. Is Homogenizing Our Thoughts," *The New Yorker*, June 2025.

Choi, Jonathan H. and Daniel Schwarcz, "AI Assistance in Legal Analysis: An Empirical Study," *Working Paper*, 2023.

Choudhury, Prithwiraj, Evan Starr, and Rajshree Agarwal, "Machine learning and human capital complementarities: Experimental evidence on bias mitigation," *Strategic Management Journal*, 2020.

Conti, Annamaria and Danilo Messinese, "The Selective Tailwind Effect of Artificial Intelligence," *Available at SSRN*, 2024.

_ and Jorge A Guzman, "What is the US comparative advantage in entrepreneurship? Evidence from Israeli migration to the United States," *Review of Economics and Statistics*, 2023, *105* (3), 528–544.

Cowgill, Bo, "Bias and productivity in humans and algorithms: Theory and evidence from resume screening.," *Working Paper Available at SSRN 3433737*, 2020.

Dawes, Robyn M, "The robust beauty of improper linear models in decision making.," *American psychologist*, 1979, *34* (7), 571.

Dell'Acqua, Fabrizio, Bruce Kogut, and Patryk Perkowski, "Super Mario Meets AI: Experimental Effects of Automation and Skills on Team Performance and Coordination," *Review of Economics and Statistics*, 2023, pp. 1–47.

__ , Edward McFowland, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani, "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 Available at SSRN 4573321*, 2023.

Dell'Acqua, Fabrizio, "Falling asleep at the wheel: Human/AI collaboration in a field experiment on HR recruiters.," *Working Paper*, 2022.

Doleac, Jennifer L and Benjamin Hansen, "The unintended consequences of "ban the box": Statistical discrimination and employment outcomes when criminal histories are hidden," *Journal of Labor Economics*, 2020, *38* (2), 321–374.

Doshi, Anil R and Oliver P Hauser, "Generative AI enhances individual creativity but reduces the collective diversity of novel content," *Science advances*, 2024, *10* (28), eadn5290.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: Labor market impact potential of LLMs," *Science*, 2024, *384* (6702), 1306–1308.

Felten, Ed, Manav Raj, and Robert Seamans, "How will Language Modelers like ChatGPT Affect Occupations and Industries?," *arXiv preprint arXiv:2303.01157*, 2023.

Filippas, Apostolos, John Joseph Horton, and Joseph Golden, "Reputation inflation," in "Proceedings of the 2018 ACM Conference on Economics and Computation" 2018, pp. 483–484.

Flesch, Rudolf, "Flesch-Kincaid readability test," *Retrieved October*, 2007, *26* (3), 2007.

Gans, Joshua S., "How will Generative AI impact communication?," *Economics Letters*, 2024, p. 111872.

__ , *The Microeconomics of Artificial Intelligence* 2024. Forthcoming.

__ , Scott Stern, and Jane Wu, "Foundations of entrepreneurial strategy," *Strategic Management Journal*, 2019.

Ghose, Anindya and Panagiotis G Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE transactions on knowledge and data engineering*, 2010, *23* (10), 1498–1512.

Gupta, Budhaditya and Tarun Khanna, "A recombination-based internationalization model: Evidence from Narayana health's journey from India to the Cayman Islands," *Organization Science*, 2019.

Habgood-Coote, Joshua, "Deepfakes and the epistemic apocalypse," *Synthese*, 2023, *201* (3), 103.

Hastie, Trevor, Robert Tibshirani, and Jerome H Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Vol. 2, Springer, 2017.

Holmström, Bengt, "Managerial incentive problems: A dynamic perspective," *The review of Economic studies*, 1999, *66* (1), 169–182.

Holmstrom, Bengt and Paul Milgrom, "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design," *The Journal of Law, Economics, and Organization*, 1991, *7* (special issue), 24–52.

Hong, Yili, Jing Peng, Gordon Burtch, and Ni Huang, "Just DM me (politely): Direct messaging, politeness, and hiring outcomes in online labor markets," *Information Systems Research*, 2021, *32* (3), 786–800.

Horton, John, "Blockchain as a Solution to Trust Issues," February 16 2024. Accessed: 2024-07-22.

Horton, John J, "The effects of algorithmic labor market recommendations: Evidence from a field experiment," *Journal of Labor Economics*, 2017, *35* (2), 345–385.

Hui, Xiang, Oren Reshef, and Luofeng Zhou, "The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market," *Organization Science*, 2023.

James, William and Charles Stein, "Estimation with Quadratic Loss," in "Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics," Vol. 4 University of California Press 1961, pp. 361–380.

Jia, Nan, Xueming Luo, Zheng Fang, and Chengcheng Liao, "When and how artificial intelligence augments employee creativity," *Academy of Management Journal*, 2024, *67* (1), 5–32.

Kerr, Steven, "On the folly of rewarding A, while hoping for B," *Academy of Management journal*, 1975, *18* (4), 769–783.

Kerr, William R and Frederic Robert-Nicoud, "Tech clusters," *Journal of Economic Perspectives*, 2020, *34* (3), 50–76.

Kim, Hyunjin, Edward L Glaeser, Andrew Hillis, Scott Duke Kominers, and Michael Luca, "Decision authority and the returns to algorithms," *Strategic Management Journal*, 2024, *45* (4), 619–648.

Kleinberg, Jon and Manish Raghavan, "Algorithmic monoculture and social welfare," *Proceedings of the National Academy of Sciences*, 2021, *118* (22), e2018340118.

Kokkodis, Marios and Sam Ransbotham, "Learning to successfully hire in online labor markets," *Management Science*, 2023, *69* (3), 1597–1614.

Kosmyna, Nataliya, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes, "Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task," *arXiv preprint arXiv:2506.08872*, 2025.

Li, Danielle, Lindsey R. Raymond, and Peter Bergman, "Hiring as exploration," *NBER Working Paper, No. w27736*, 2020.

List, John A, "Non est disputandum de generalizability? A glimpse into the external validity trial," Technical Report, National Bureau of Economic Research 2020.

Marinescu, Ioana and Ronald Wolthoff, "Opening the black box of the matching function: The power of words," *Journal of Labor Economics*, 2020, *38* (2), 535–568.

Martin-Lacroux, Christelle and Alain Lacroux, "Do employers forgive applicants' bad spelling in résumés?," *Business and Professional Communication Quarterly*, 2017, *80* (3), 321–335.

McDaniel, Michael A, Deborah L Whetzel, Frank L Schmidt, and Steven D Maurer, "The validity of employment interviews: A comprehensive review and meta-analysis.," *Journal of applied psychology*, 1994, *79* (4), 599.

McElheran, Kristina, J Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia Foster, and Nikolas Zolas, "AI adoption in America: Who, what, and where," *Journal of Economics & Management Strategy*, 2024, *33* (2), 375–415.

Nagaraj, Abhishek, Esther Shears, and Mathijs de Vaan, "Improving data access democratizes and diversifies science," *Proceedings of the National Academy of Sciences*, 2020, *117* (38), 23490–23498.

Nanda, Ramana, Sampsa Samila, and Olav Sorenson, "The persistent effect of initial success: Evidence from venture capital," *Journal of Financial Economics*, 2020, *137* (1), 231–248.

Nosko, Chris and Steven Tadelis, "The limits of reputation in platform markets: An empirical analysis and field experiment," Technical Report, National Bureau of Economic Research 2015.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 2023, *381* (6654), 187–192.

Otis, Nicholas G, Katelyn Cranney, Solene Delecourt, and Rembrand Koning, "Global Evidence on Gender Gaps and Generative AI," Technical Report, Center for Open Science 2024.

Otis, Nicholas G., Rowan Clarke, Solene Delecourt, David Holtz, and Rembrand Koning, "The Uneven Impact of Generative AI on Entrepreneurial Performance," *Working Paper*, 2023.

Pritchard, Jonathan K, Matthew Stephens, and Peter Donnelly, "Inference of population structure using multilocus genotype data," *Genetics*, 2000, *155* (2), 945–959.

Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever et al., "Language models are unsupervised multitask learners," *OpenAI blog*, 2019, *1* (8), 9.

Raghavan, Manish, "Competition and diversity in generative ai," *arXiv preprint arXiv:2412.08610*, 2024.

Sackett, Paul R, Charlene Zhang, Christopher M Berry, and Filip Lievens, "Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range.," *Journal of Applied Psychology*, 2022, *107* (11), 2040.

Sajjadiani, Sima, Aaron J Sojourner, John D Kammeyer-Mueller, and Elton Mykerezi, "Using machine learning to translate applicant work history into predictors of performance and turnover.," *Journal of Applied Psychology*, 2019, *104* (10), 1207.

Sawy, Omar El, Milan Miric, and Margherita Pagani, "Digital platform ecosystems: the coming context for AI," in "Artificial Intelligence for Sustainable Value Creation," Edward Elgar Publishing, 2021, pp. 55–69.

Schmidt, Frank L and John E Hunter, "The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings.," *Psychological bulletin*, 1998, *124* (2), 262.

Scott, Erin L., Pian Shu, and Roman M. Lubynsky, "Entrepreneurial uncertainty and expert evaluation: An empirical analysis," *Management Science*, 2020.

Spence, Michael, "Job Market Signaling," *The Quarterly Journal of Economics*, 1973, *87* (3), 355–374.

Stein, Charles, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in "Proceedings of the Third Berkeley symposium on mathematical statistics and probability," Vol. 1 1956, pp. 197–206.

Sterkens, Philippe, Ralf Caers, Marijke De Couck, Victor Van Driessche, Michael Geamanu, and Stijn Baert, "Costly mistakes: Why and when spelling errors in resumes jeopardise interview chances," *Plos one*, 2023, *18* (4), e0283280.

Svanberg, Maja, Wensu Li, Martin Fleming, Brian Goehring, and Neil Thompson, "Beyond AI Exposure: Which Tasks are Cost-Effective to Automate with Computer Vision?," *Available at SSRN 4700751*, 2024.

Tambe, Prasanna, "Big data investment, skills, and firm value," *Management science*, 2014, *60* (6), 1452–1469.

Toups, Connor, Rishi Bommasani, Kathleen Creel, Sarah Bana, Dan Jurafsky, and Percy S Liang, "Ecosystem-level analysis of deployed machine learning reveals homogeneous outcomes," *Advances in Neural Information Processing Systems*, 2023, *36*, 51178–51201.

Weiser, Benjamin, "Here's What Happens When Your Lawyer Uses ChatGPT," *New York Times*, 2023. Available at: https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html.

Weiss, Daphne, Sunny X Liu, Hannah Mieczkowski, and Jeffrey T Hancock, "Effects of using artificial intelligence on interpersonal perceptions of job applicants," *Cyberpsychology, Behavior, and Social Networking*, 2022, *25* (3), 163–168.

Wiles, Emma, Zanele Munyikwa, and John J. Horton, "The Impact of AI Writing Assistance on Job Posts and the Supply of Jobs on an Online Labor Market," *Working Paper*, 2023.

_ , Zanele T. Munyikwa, and John J. Horton, "Algorithmic writing assistance on jobseekers' resumes increases hires," *NBER Working Paper No. w30886*, 2023.

Wright, Nataliya L., "When local learning scales: Entrepreneurs' initial users and market expansion," *Organization Science*, 2025, (Accepted).

_ , Frank Nagle, and Shane Greenstein, "Open source software and global entrepreneurship," *Research Policy*, 2023.

_ , Rembrand Koning, and Tarun Khanna, "Judging foreign startups," *Strategic Management Journal*, 2023.

Appendix: For Online Publication Only

Table of Contents

A Theory Appendix

Table A1: **Expanded Glossary of Notation used in Results/Proofs**

Symbol	**Meaning / Definition**	**Support / Type**
Beliefs, payoffs, and equilibrium objects		
$\mathbb{E}[e_n \mid s, G]$	Receiver's posterior mean belief about expertise.	$\mathbb{R}$
$\pi(G)$	Weight on the signal in Bayesian updating.	[0,1]
$P(s, G) = r \cdot \mathbb{E}[e_n \mid s, G]$	Payment to sender.	$\mathbb{R}$
r	Sender's revenue share.	$\mathbb{R}_{>0}$
$u_s = P - \frac{a^2}{2}$	Sender's utility.	$\mathbb{R}$
Accuracy and loss		
$f_B(b; x, G)$	PDF of receiver beliefs b given true expertise x.	density
$L(x, G)$	Squared-error loss $(x - b)^2$ averaged over beliefs.	$\mathbb{R}_{\geq 0}$
$\rho(G)$	Correlation between e_n and signal under G.	[-1,1]
$\mathbb{E}_x[L(x, G)]$	Average loss $(1 - \rho(G)^2)\sigma_n^2$.	$\mathbb{R}_{\geq 0}$
Comparative statics and thresholds		
σ^-, σ^+	Cutoffs: accuracy improves if $\sigma_{ng} \notin (\sigma^-, \sigma^+)$.	constants
σ^S	Cutoff for mean signal increase ($\sigma^S < 0$).	constant
$\text{Var}(s \mid G)$	Variance of signals in population.	$\mathbb{R}_{\geq 0}$
Dispersion cond.	$\text{Var}(s \mid 1) > \text{Var}(s \mid 0) \iff \sigma_{ng}/\sigma_g^2 \geq -1/2$.	condition

Notes. This table summarizes notation in the setup of our model. Table 1 in the main text is a similar table about the basic setup.

Lemma A1. *The receiver's (mean) beliefs about the sender's expertise are:*

$$\mathbb{E}[e_n \mid s, G] = \bar{e}_n + \pi(G)[s - a^* - \bar{e}_n - G \cdot \bar{e}_g], \tag{4}$$

where $\pi(G)$, the weight placed on the signal, equals

$$\pi(G) = \frac{Cov(e_n, s - a^*)}{Var(s - a^*)} = \frac{\sigma_n^2 + G\sigma_{ng}}{\sigma_n^2 + 2G\sigma_{ng} + G^2\sigma_g^2 + \sigma_\epsilon^2}, \tag{5}$$

and optimal effort a^* is

$$a^* = r\pi(G). \tag{6}$$

Proof of Lemma A1. The expression uses the normality assumptions of the problem; namely, the conditional expectation is a weighted sum of the prior and the signal. Following Holmström (1999), the receiver infers optimal effort by the sender (a^*), so even though the sender's effort is not observable, it can be used by the receiver to correct the signal.

Because the receiver does not directly observe the sender's effort, the sender might seek to manipulate the receiver's beliefs. Anticipating this, the receiver knows that the sender chooses effort a to maximize Equation 2, so the sender's optimal effort a^* satisfies the first order condition $a^* = r \cdot \partial \mathbb{E}[e_n|s, G]/\partial a$. The derivative of the payment with respect to a captures the returns the sender expects to get from manipulating the receiver's beliefs. The receiver is aware of this manipulation motive, and even though the sender's effort is unobservable, the receiver can infer it from solving Equation 2 (e.g., Holmström 1999). As a result, observing the signal s is equivalent to observing

$$s - a^* = e_n + G \cdot e_g + \epsilon.$$

Bénabou and Tirole (2006, Proposition 1) show that the optimal effort does not depend on the types, and a unique closed-form solution for the effort can be found in (a differential-reputation) equilibrium: $a^* = r\pi(G)$, where

$$\pi(G) = \frac{\sigma_n^2 + G\sigma_{ng}}{\sigma_n^2 + 2G\sigma_{ng} + G^2\sigma_g^2 + \sigma_\epsilon^2}.$$

□

Proof of Corollary 1. From Equation 5, without GAI

$$\pi(0) = \frac{\sigma_n^2}{\sigma_n^2 + \sigma_\epsilon^2}$$

and with GAI

$$\pi(1) = \frac{\sigma_n^2 + \sigma_{ng}}{\sigma_n^2 + 2\sigma_{ng} + \sigma_g^2 + \sigma_\epsilon^2}.$$

Comparing both expressions yields that GAI increases the weight, namely, $\pi(1) > \pi(0)$, if and only if

- $\sigma_{ng} > \frac{\sigma_n^2\sigma_g^2}{\sigma_\epsilon^2 - \sigma_n^2}$ when $\sigma_\epsilon^2 > \sigma_n^2$, or
- $\sigma_{ng} < -\frac{\sigma_n^2\sigma_g^2}{\sigma_n^2 - \sigma_\epsilon^2}$ when $\sigma_\epsilon^2 < \sigma_n^2$,

and decreases it otherwise. The effect of GAI can go in either direction because there is no restriction on the parameter values other than that the variances are strictly positive and that the covariance σ_{ng} is in the interval $[-\sigma_n\sigma_g, \sigma_n\sigma_g]$. For example, consider σ_ϵ arbitrarily close to zero. If σ_{ng} is sufficiently below $-\sigma_g^2$, GAI increases the weight. If σ_{ng} is greater than $-\sigma_g^2$, GAI decreases it. □

Proof of Corollary 2. Optimal effort in equilibrium is $a^* = r\pi(G)$ (Equation 6), where $r > 0$ constant, so the effect of GAI on effort is proportional to the effect of GAI on the weight $\pi(G)$. The effect of GAI on the weight follows from Corollary 1, which can go in either direction. □

Lemma A2 (Information Loss)**.** *Loss $L(x,G)$ is given by:*

$$L(x,G) = \left(1-\rho(G)^2\right)\left(\left(1-\rho(G)^2\right)(x-\bar{e}_n)^2 + \rho(G)^2\sigma_n^2\right), \tag{7}$$

where $\rho(G)$ is the correlation between expertise and the signal in equilibrium, which is

$$\rho(G) \equiv \frac{\sigma_n^2 + G\sigma_{ng}}{\sigma_n\sqrt{\sigma_n^2 + 2G\sigma_{ng} + G^2\sigma_g^2 + \sigma_\epsilon^2}}.$$

(This result shows that receivers' loss when evaluating a sender of expertise $e_n = x$ is proportional to a weighted sum of its squared deviation from average expertise and the variance of expertise.)

Proof of Lemma A2. The Loss function measures the receiver's beliefs' expected deviations when the sender's true expertise is $e_n = x$:

$$\begin{aligned}
L(x,G) &= \int (x-b)^2 f_B(x,G)(b)db \\
&= \int (x^2 - 2bx + b^2) f_B(x,G)(b)db \\
&= x^2 - 2x\int b f_B(x,G)(b)db + \int b^2 f_B(x,G)(b)db \\
&= x^2 - 2xE[\mathcal{B}|e_n = x, G] + E[\mathcal{B}^2|e_n = x, G]
\end{aligned}$$

The first line is the definition. The second line comes from expanding the squared function, the third from integrating each summand, and the final line from defining $\mathcal{B} \equiv E[e_n|s-a^*, G]$, which is a random variable that represents the receivers' beliefs.

Let us now find expressions for both $E[\mathcal{B}|e_n = x, G]$ and $E[\mathcal{B}^2|e_n = x, G]$ as functions of the primitive parameters. Denote $f_{e_g,\epsilon|e_n=x}(e_g,\epsilon)$ the joint distribution of (e_g,ϵ) conditional on $e_n = x$. Let us start by computing $E[\mathcal{B}|e_n = x, G]$:

$$
\begin{aligned}
E[\mathcal{B}|e_n = x, G] &= E[\underbrace{E[e_n|s-a^*, G)]}_{\text{Receivers' beliefs}}|e_n = x, G] \\
&= \iint E[e_n|s-a^*, G)] f_{e_g,\epsilon|e_n=x}(e_g,\epsilon) de_g d_\epsilon \\
&= \iint (\bar{e}_n + \pi(G)(s-a^* - E(s-a^*)))) f_{e_g,\epsilon|e_n=x}(e_g,\epsilon) de_g d_\epsilon \\
&= \bar{e}_n + \pi(G)\left(x + \iint (Ge_g + \epsilon) f_{e_g,\epsilon|e_n=x}(e_g,\epsilon) de_g d_\epsilon - \bar{e}_n - G\bar{e}_g\right) \\
&= \bar{e}_n + \pi(G)\left(x + E(Ge_g + \epsilon|e_n = x) - \bar{e}_n - G\bar{e}_g\right) \\
&= \bar{e}_n + \pi(G)\left(x + G\bar{e}_g + \frac{G\sigma_{ng}}{\sigma_n^2}(x - \bar{e}_n) - \bar{e}_n - G\bar{e}_g\right) \\
&= \bar{e}_n + \pi(G)\left(1 + \frac{G\sigma_{ng}}{\sigma_n^2}\right)(x - \bar{e}_n) \\
&= \bar{e}_n + \rho(G)^2(x - \bar{e}_n),
\end{aligned}
$$

where the second equality comes from the definition of conditional expectation; the third line from Equation 5 in Lemma A1; the fourth line from the fact that $s - a^* = x + Ge_g + \epsilon$ when $e_n = x$ and a^* does not depend on either e_n, e_g, or ϵ; the fifth line from the definition of conditional expectation; the sixth line from the normality assumptions of the problem and linearity; the seventh and eighth lines from arranging terms.

Lastly, applying the definition of the correlation coefficient yields:

$$
\rho(G)^2 = \frac{Cov(e_n, s-a^*)^2}{Var(e_n)Var(s-a^*)} = \frac{(\sigma_n^2 + G\sigma_{ng})^2}{\sigma_n^2(\sigma_n^2 + 2G\sigma_{ng} + G^2\sigma_g^2 + \sigma_\epsilon^2)}.
$$

To compute $E[\mathcal{B}^2|e_n = x, G]$, note that $Var[Z] = E[Z^2] - E[Z]^2$ for any random variable Z with finite variance. Thus,

$$
\begin{aligned}
E[\mathcal{B}^2|e_n = x, G] &= Var[\mathcal{B}|e_n = x, G] + E[\mathcal{B}|e_n = x, G]^2 \\
&= \left(1 - \frac{Cov(\mathcal{B}, e_n)^2}{Var(\mathcal{B})Var(e_n)}\right) Var(\mathcal{B}) + E[\mathcal{B}|e_n = x, G]^2,
\end{aligned}
$$

The first term takes that form because $\mathcal{B}$ is normally distributed as a linear function of normally distributed random variables. In other words, this means that knowing that expertise $e_n = x$ amounts to reducing the variance of receivers' beliefs by a factor of $\left(1 - \frac{Cov(\mathcal{B},e_n)^2}{Var(\mathcal{B})Var(e_n)}\right)$, which becomes smaller the more extreme is the correlation between beliefs and expertise: either very positive (close to one) or very negative (close to zero). The second term is just $\left(\bar{e}_n + \rho(G)^2(x - \bar{e}_n)\right)^2$ as derived above. Replacing with the distributional parameters:

$$
\begin{aligned}
E[\mathcal{B}^2|e_n = x, G] &= \left(1 - \frac{(\pi(G)(\sigma_n^2 + G\sigma_{ng}))^2}{\pi(G)^2(\sigma_n^2 + 2G\sigma_{ng} + G^2\sigma_g^2 + \sigma_\epsilon^2)\sigma_n^2}\right)\pi(G)^2(\sigma_n^2 + 2G\sigma_{ng} + G^2\sigma_g^2 + \sigma_\epsilon^2) \\
&\quad + \left(\bar{e}_n + \rho(G)^2(x - \bar{e}_n)\right)^2 \\
&= \left(1 - \rho(G)^2\right)\rho(G)^2\sigma_n^2 + \left(\bar{e}_n + \rho(G)^2(x - \bar{e}_n)\right)^2.
\end{aligned}
$$

Plugging in the expressions just derived in the original equation, the Loss function can be written as

$$
\begin{aligned}
L(x, G) &= x^2 - 2x\left(\bar{e}_n + \rho(G)^2(x - \bar{e}_n)\right) + \left(1 - \rho(G)^2\right)\rho(G)^2\sigma_n^2 + \left(\bar{e}_n + \rho(G)^2(x - \bar{e}_n)\right)^2 \\
&= \left(x - \bar{e}_n - \rho(G)^2(x - \bar{e}_n)\right)^2 + \left(1 - \rho(G)^2\right)\rho(G)^2\sigma_n^2 \\
&= \left(1 - \rho(G)^2\right)^2(x - \bar{e}_n)^2 + \left(1 - \rho(G)^2\right)\rho(G)^2\sigma_n^2 \\
&= \left(1 - \rho(G)^2\right)\left(\left(1 - \rho(G)^2\right)(x - \bar{e}_n)^2 + \rho(G)^2\sigma_n^2\right).
\end{aligned}
$$

The Loss function is quadratic in x, and it is straightforward to check that the minimum point is $x = \bar{e}_n$ and the minimum value $\left(1 - \rho(G)^2\right)\rho(G)^2\sigma_n^2 > 0$. □

Figure A1 illustrates the Loss function as a function of expertise x, when GAI is and is not available.

Proof of Lemma 1. The result follows from taking expectations over x (expertise) on Equation 7. □

Proof of Proposition 1. From the expressions of expected accuracy loss,

$$
\mathbb{E}_x[L(x, 1)] < \mathbb{E}_x[L(x, 0)] \iff \rho(1)^2 > \rho(0)^2.
$$

In words: GAI availability reduces expected loss if and only if the squared correlation coefficient between expertise and the signal (adjusted by effort) is greater than the squared correlation coefficient when GAI is not available.

Let us now compare those squared correlation coefficients. The expressions for the squared correlations are:

$$
\left(\frac{\sigma_n^2 + \sigma_{ng}}{\sigma_n\sqrt{\sigma_n^2 + 2\sigma_{ng} + \sigma_g + \sigma_\epsilon^2}}\right)^2 > \left(\frac{\sigma_n^2}{\sigma_n\sqrt{\sigma_n^2 + \sigma_\epsilon^2}}\right)^2,
$$

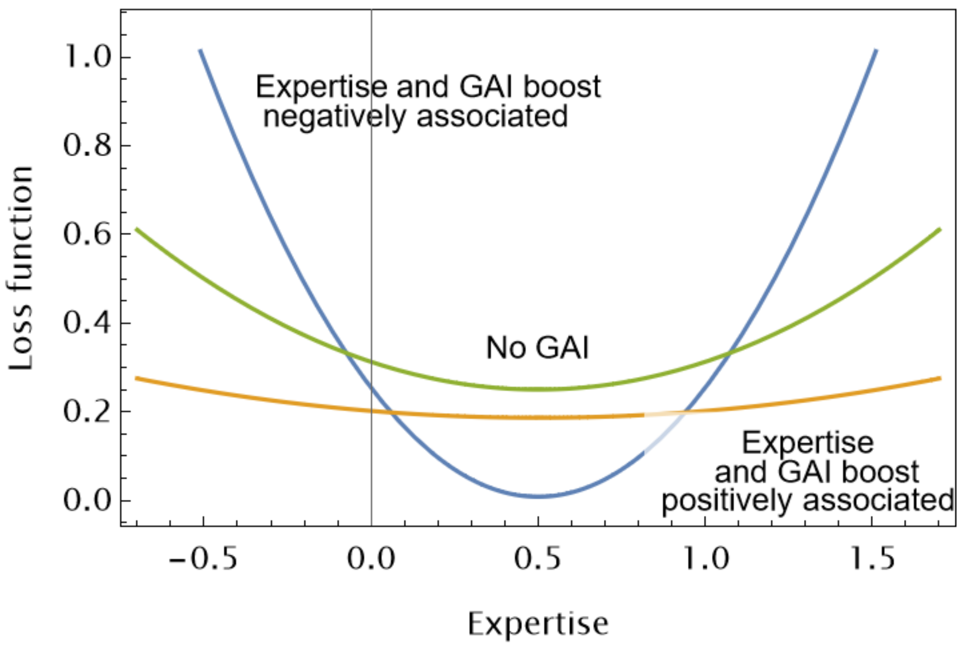


Figure A1: Loss for different covariance values between expertise and GAI boost. All other noise parameters are normalized to one, and the mean expertise is set to 0.5. For example, when the covariance between expertise and GAI boost is negative (set to -0.9 in the figure), the loss of screening extreme types is larger than when the covariance is positive (0.9); but the loss is smaller when screening intermediate types. Another observation is that GAI does not necessarily make screening less accurate. For example, GAI dominates no GAI when the covariance between GAI boost and expertise is positive.

which after straightforward algebra yields:

$$\iff \left(\sigma_{ng} + \frac{\sigma_n^2\sigma_\epsilon^2}{\sigma_n^2+\sigma_\epsilon^2}\right)^2 > \frac{(\sigma_n^2)^2\sigma_g^2}{\sigma_n^2+\sigma_\epsilon^2} + \left(\frac{\sigma_n^2\sigma_\epsilon^2}{\sigma_n^2+\sigma_\epsilon^2}\right)^2,$$

which leads to two conditions:

$$\sigma_{ng} > \sqrt{\frac{(\sigma_n^2)^2\sigma_g^2}{\sigma_n^2+\sigma_\epsilon^2} + \left(\frac{\sigma_n^2\sigma_\epsilon^2}{\sigma_n^2+\sigma_\epsilon^2}\right)^2} - \frac{\sigma_n^2\sigma_\epsilon^2}{\sigma_n^2+\sigma_\epsilon^2}$$

or

$$\sigma_{ng} < -\sqrt{\frac{(\sigma_n^2)^2\sigma_g^2}{\sigma_n^2+\sigma_\epsilon^2} + \left(\frac{\sigma_n^2\sigma_\epsilon^2}{\sigma_n^2+\sigma_\epsilon^2}\right)^2} - \frac{\sigma_n^2\sigma_\epsilon^2}{\sigma_n^2+\sigma_\epsilon^2}.$$

Simplifying further, the first inequality leads to:

$$\sigma_{ng} > \sigma^+ \equiv \frac{\sigma_n^2}{\sigma_n^2+\sigma_\epsilon^2}\left(\sqrt{(\sigma_\epsilon^2)^2 + \sigma_g^2(\sigma_n^2+\sigma_\epsilon^2)} - \sigma_\epsilon^2\right) > 0.$$

And the second inequality yields a negative cutoff:

$$\sigma_{ng} < \sigma^- \equiv -\frac{\sigma_n^2}{\sigma_n^2+\sigma_\epsilon^2}\left(\sqrt{(\sigma_\epsilon^2)^2 + \sigma_g^2(\sigma_n^2+\sigma_\epsilon^2)} + \sigma_\epsilon^2\right) < 0.$$

□

Proof of Proposition 2. Define $s^*(G) = a^*(G) + e_n + Ge_g + \epsilon$, the level of the signal in equilibrium for GAI availability $G = 0, 1$. We need to establish that $\mathbb{E}[s^*(1) - s^*(0)] \geq 0$, which is equivalent to $\pi(1) - \pi(0) \geq -\bar{e}_g/r$ if and only if σ_{ng} is greater or equal than a negative constant cutoff.

Plugging in the expressions for $\pi(1)$ and $\pi(0)$, the cutoff comes from straightforward algebra,

$$\sigma_{ng} \geq \sigma^S \equiv -\frac{\bar{e}_g(\sigma_n^2+\sigma_g^2+\sigma_\epsilon^2)(\sigma_n^2+\sigma_\epsilon^2) - r\sigma_n^2\sigma_g^2}{\sigma_n^2(2\bar{e}_g - r) + \sigma_\epsilon^2(2\bar{e}_g + r)}.$$

Note that σ^S is a negative constant because the numerator $\bar{e}_g(\sigma_n^2+\sigma_g^2+\sigma_\epsilon^2)(\sigma_n^2+\sigma_\epsilon^2) - r\sigma_n^2\sigma_g^2$ and the denominator $\sigma_n^2(2\bar{e}_g - r) + \sigma_\epsilon^2(2\bar{e}_g + r)$ are both positive under the assumptions. □

Proof of Proposition 3. Defining $s^*(G) = a^*(G) + e_n + Ge_g + \epsilon$, the level of the signal in equilibrium for GAI availability $G = 0, 1$, the result comes directly from comparing the variance of the signal with GAI $Var(s^*(1)) = \sigma_n^2 + 2\sigma_{ng} + \sigma_g^2 + \sigma_\epsilon^2$ and $Var(s^*(0)) = \sigma_n^2 + \sigma_\epsilon^2$. □

A1 Receivers' beliefs

Generative AI largely affects the signal and its variance by incentivizing the sender to manipulate the receiver. However, a rational receiver takes that manipulation motive into account when making inferences about expertise, as shown in Equation 4. The signal jamming channel through which GAI affects the sender's signal does not affect the receiver's beliefs.

Rather, the effect of GAI on the receiver's beliefs manifests primarily in the weight $\pi(G)$ in Equation 5 the receiver puts on the signal. We call this weight the "informativeness" of the signal. Again, the covariance plays an important role in determining the effect of GAI on how informative the signal is:

Lemma A3. *Let $\pi(G) = (\sigma_n^2 + G\sigma_{ng})/(\sigma_n^2 + 2G\sigma_{ng} + G^2\sigma_g^2 + \sigma_\epsilon^2)$ be the weight the receiver puts on the signal (net of the signal's mean) as a function of GAI availability $G = 0, 1$. The receiver puts more weight on the signal when GAI is available if and only if*

- $\sigma_{ng} > \frac{\sigma_n^2\sigma_g^2}{\sigma_\epsilon^2 - \sigma_n^2}$ *when* $\sigma_\epsilon^2 > \sigma_n^2$, *or*
- $\sigma_{ng} < -\frac{\sigma_n^2\sigma_g^2}{\sigma_n^2 - \sigma_\epsilon^2}$ *when* $\sigma_\epsilon^2 < \sigma_n^2$.

Proof of Lemma A3. The result follows from comparing $\pi(1) = (\sigma_n^2 + \sigma_{ng})/(\sigma_n^2 + 2\sigma_{ng} + \sigma_g^2 + \sigma_\epsilon^2)$ and $\pi(0) = \sigma_n^2/(\sigma_n^2 + \sigma_\epsilon^2)$ such that $\pi(1) > \pi(0)$. $\square$

The last result highlights the difficulties in assessing expertise when GAI is available. GAI availability can be *more* informative than no GAI availability for plausible parameter values. For example, if the task is subject to large idiosyncratic variation, for example $\sigma_\epsilon^2 > \sigma_n^2$, a large positive covariance between expertise and GAI boost makes the receiver put more weight on the (noisy) signal when GAI is available than when GAI is not available. In contrast, when the signal perfectly reveals the sender's expertise without GAI, $\sigma_\epsilon^2 = 0$, a negative, large-magnitude covariance makes GAI availability more informative than no GAI availability. A corollary of Proposition A3 is that a covariance close enough to zero makes the signal less informative when GAI is available for any parameter value.

Corollary 3. *A sufficient condition for GAI to make the signal less informative is*

$$\sigma_{ng} \in I_\pi,$$

where

$$I_\pi \equiv \left(-\frac{\sigma_n^2\sigma_g^2}{\max\{\sigma_\epsilon^2, \sigma_n^2\} - \min\{\sigma_\epsilon^2, \sigma_n^2\}}, \frac{\sigma_n^2\sigma_g^2}{\max\{\sigma_\epsilon^2, \sigma_n^2\} - \min\{\sigma_\epsilon^2, \sigma_n^2\}}\right)$$

is a non-empty set centered at zero.

Corollary 3 gives precise boundaries for the intermediate values of the covariance between expertise and GAI boost that make GAI less informative. This result is useful empirically because values of the covariance around zero indicate that GAI reduces the informativeness of the signal. Our experimental design allows us to identify the covariance following the identification result in Proposition B1.

A2 The receiver does not observe the sender's GAI use.

Let GAI be available but its use is denoted by G, a continuous random variable. Assume also that e_g is fixed at one. The key assumption is that GAI use G inherits the distributional parameters of e_g in the original model, namely the mean of G is $\bar{e}_g$, its variance is σ_g^2, and covariance with e_n is σ_{ng}. The assumption on the continuity of G has the benefit of tractability but the cost of implausible negative values. As in Bénabou and Tirole (2006, p.1660), however, we assume the relevant means are high enough to make the probability of the negative values arbitrarily small. We also adopt their interpretation of the following results as local approximations. An evaluator's beliefs about the sender's GAI use is

$$\begin{aligned}\mathbb{E}[G|s] &= \bar{e}_g + \frac{Cov(G,s)}{Var(s)}(s - a^* - \bar{e}_n - \bar{e}_g) \\ &= \bar{e}_g + \frac{\sigma_g^2 + \sigma_{ng}}{\sigma_n^2 + 2\sigma_{ng} + \sigma_g^2 + \sigma_\epsilon}(s - a^* - \bar{e}_n - \bar{e}_g).\end{aligned} \tag{8}$$

with a^* defined now as

$$a^* = r\frac{\sigma_n^2 + \sigma_{ng}}{\sigma_n^2 + 2\sigma_{ng} + \sigma_g^2 + \sigma_\epsilon}$$

and $s = e_n + G + a^* + \epsilon$ in equilibrium.

Equation 8 allows us to clarify two important points. First, the model provides a prescription for GAI use after re-interpreting GAI boost as GAI use. The normality assumption is for exposition only, as it makes the updating less cumbersome.

The second clarification is the importance of the covariance between expertise and GAI use. A receiver's beliefs about GAI use depend on this covariance in a similar way a receiver's beliefs about expertise do. The lesson from this exercise is that our model also prescribes the evaluator's inferences about GAI use without changing anything fundamental in the model.

In the paper, we do not emphasize the evaluator's inferences about GAI use because our experiment is designed to measure accuracy when screening expertise, rather than GAI use. A different paper could ask that question: What determines GAI use when it is not observable by an evaluator? We agree that this could be an important question for future experimental work, for which our model provides insights. For example, Equation 8 provides hypotheses for GAI adoption that might be tested in future work. For another example, the effect of exogenous variation in any of the variance-covariance parameters on GAI adoption comes directly from Equation 8.

A3 Endogenous GAI Use

With a proper re-interpretation of the parameters, our model can also shed light on the receivers' beliefs about the sender's expertise and GAI use when the latter is the sender's choice.

Consider the model in the paper again. Assume that GAI is available and let us re-interpret effort a as GAI use intensity. As such, the receiver does not observe the sender's GAI use. Because the receiver infers GAI use even if it is not observed as demonstrated by Holmström (1999, p.171 eq.5) under similar assumptions, the problem of inferring GAI use would not be interesting. Instead, we assume that GAI use depends on the sender's type. We could model this type in many ways. The one that requires fewer additional assumptions is to assume that the sender's return on beliefs r is now a normal random variable, with mean $\bar{r}$ and variance σ_r^2. To maintain parsimony, we assume that r is correlated with expertise, the covariance is σ_{nr}, and r is not correlated with the other random variables.

The re-interpretation of private effort as GAI adoption and assumption of the sender's payoff depending on their type are sufficient to allow the model to yield prescriptions for the receiver's beliefs about GAI use in a context where adoption is endogenous. To keep the notation to a minimum, let us assume that GAI boost e_g is a constant equal to zero. Assuming it is normally distributed does not change the qualitative results we discuss here.

The sender chooses GAI use a^* according to the first-order condition from Equation 2

$$a^* = r\frac{\partial \mathbb{E}[e_n|s]}{\partial a}.$$

We conjecture that $\partial \mathbb{E}[e_n|s]/\partial a$ does not depend on a and then corroborate that is the case. To compute the derivative, the receiver's beliefs about expertise are given by

$$\mathbb{E}[e_n|s] = \bar{e}_n + \frac{\sigma_n^2 + \sigma_{nr}\frac{\partial \mathbb{E}[e_n|s]}{\partial a}}{\sigma_n^2 + 2\sigma_{nr}\frac{\partial \mathbb{E}[e_n|s]}{\partial a} + \left(\frac{\partial \mathbb{E}[e_n|s]}{\partial a}\right)^2 \sigma_r^2 + \sigma_\epsilon^2}(s - \bar{s})$$

where s is the signal and equals $e_n + a^* + \epsilon$. If $\partial \mathbb{E}[e_n|s]/\partial a$ does not depend on a, then

$$\frac{\partial \mathbb{E}[e_n|s]}{\partial a} = \frac{\sigma_n^2 + \sigma_{nr}\frac{\partial \mathbb{E}[e_n|s]}{\partial a}}{\sigma_n^2 + 2\sigma_{nr}\frac{\partial \mathbb{E}[e_n|s]}{\partial a} + \left(\frac{\partial \mathbb{E}[e_n|s]}{\partial a}\right)^2 \sigma_r^2 + \sigma_\epsilon^2}. \tag{9}$$

Defining $\pi \equiv \partial \mathbb{E}[e_n|s]/\partial a$ and noting that $\pi = \pi(\sigma_{nr}, \sigma_n^2, \sigma_r^2, \sigma_\epsilon^2)$ there is at least one real solution π^* to Equation 9 that does not depend on a, corroborating our assumption. The solution, however, involves a non-linear function of the variance-covariance parameters. Under some assumptions, π^* can be simplified. This is left for future work.

Because the distributional assumptions in this model are the same as in Bénabou and Tirole (2006, p.1664), we make the same disclaimer: We cannot rule out the existence of equilibria that involve $\partial\mathbb{E}[e_n|s]/\partial a$ depending on a. But focusing on the solution where π^* does not depend on a fulfills the purpose of inferring expertise from the quality of the signal (pitch).

In particular, the receiver's beliefs about the sender's expertise and their GAI use are given respectively by:

$$\mathbb{E}[e_n|s] = \bar{e}_n + \frac{\sigma_n^2 + \sigma_{nr}\pi^*}{\sigma_n^2 + 2\sigma_{nr}\pi^* + (\pi^*)^2\sigma_r^2 + \sigma_\epsilon^2}(s - \bar{s})$$

and

$$\mathbb{E}[a^*|s] = \bar{r}\pi^* + \frac{(\pi^*)^2\sigma_r^2 + \sigma_{nr}\pi^*}{\sigma_n^2 + 2\sigma_{nr}\pi^* + (\pi^*)^2\sigma_r^2 + \sigma_\epsilon^2}(s - \bar{s}). \tag{10}$$

A4 Model with Binary Types

Assume no effort, and the signal is the sum of expertise and noise when GAI is *not* available ($G = 0 : s = e_n + \epsilon$) and the sum of expertise, GAI boost, and noise when GAI *is* available ($G = 1 : s = e_n + e_g + \epsilon$). These parameters follow the joint distribution given by:

Expertise	GAI boost	Noise	Probability	Signal	Signal
e_n	e_g	ϵ		$G=0$	$G=1$
0	0	0	p_{000}	0	0
0	0	1	p_{001}	1	1
0	1	0	p_{010}	0	1
0	1	1	p_{011}	1	2
1	0	0	p_{100}	1	1
1	0	1	p_{101}	2	2
1	1	0	p_{110}	1	2
1	1	1	p_{111}	2	3

The signal s equals $s = e_n + e_g G + \epsilon$ where G is GAI availability.

When GAI is not available, the signal leads to the below updating, expected receiver's beliefs conditional on the sender's true expertise, and accuracy loss:

Signal s		Beliefs b	$P[b\|e_n = 0, G = 0]$	$P[b\|e_n = 1, G = 0]$
		$b = P(e_n = 1\|s)$		
0		0	$\frac{p_{000}+p_{010}}{p_{000}+p_{001}+p_{010}+p_{011}}$	0
1		$\frac{p_{100}+p_{110}}{p_{001}+p_{011}+p_{100}+p_{110}}$	$\frac{p_{001}+p_{011}}{p_{000}+p_{001}+p_{010}+p_{011}}$	$\frac{p_{100}+p_{110}}{p_{100}+p_{101}+p_{110}+p_{111}}$
2		1	0	$\frac{p_{101}+p_{111}}{p_{100}+p_{101}+p_{110}+p_{111}}$

Beliefs b	Accuracy Loss when $e_n = 0$	Accuracy Loss when $e_n = 1$
$b = P(e_n = 1\|s)$	$L(0,0) = \sum_b (0-b)^2 P[b\|e_n = 0, G = 0]$	$L(1,0) = \sum_b (1-b)^2 P[b\|e_n = 1, G = 0]$
0	$\left(\frac{p_{100}+p_{110}}{p_{001}+p_{011}+p_{100}+p_{110}}\right)^2 \times \frac{p_{001}+p_{011}}{p_{000}+p_{001}+p_{010}+p_{011}}$	$\left(\frac{p_{001}+p_{011}}{p_{001}+p_{011}+p_{100}+p_{110}}\right)^2 \times \frac{p_{100}+p_{110}}{p_{100}+p_{101}+p_{110}+p_{111}}$
$\frac{p_{100}+p_{110}}{p_{001}+p_{011}+p_{100}+p_{110}}$		
1		

We include here the calculation of the Accuracy Loss when true expertise $e_n = 0$ for transparency:

$$
\begin{aligned}
L(e_n = 0, G = 0) = &\\
&(0-0)^2 \frac{p_{000}+p_{010}}{p_{000}+p_{001}+p_{010}+p_{011}}\\
&+ \left(0 - \frac{p_{100}+p_{110}}{p_{001}+p_{011}+p_{100}+p_{110}}\right)^2 \frac{p_{001}+p_{011}}{p_{000}+p_{001}+p_{010}+p_{011}}\\
&+ (0-1)^2 \times 0,
\end{aligned}
$$

all the other calculations are done the same way using the data from the table. Note that without GAI availability, the signal does not reveal expertise. When the signal takes the value of one, the receiver is unsure whether the sender is an expert because the pitch could have been the result of noise (luck). This reflects the world without ChatGPT, for example.

Let us compute the same statistics for when GAI is available. Note that the signal can take on one more value and that we have split the table in two because the expression are much more convoluted:

Signal s		Beliefs b	$P[b\|e_n = 0, G = 1]$	$P[b\|e_n = 1, G = 1]$
		$b = P(e_n = 1\|s)$		
0		0	$\frac{p_{000}}{p_{000}+p_{001}+p_{010}+p_{011}}$	0
1		$\frac{p_{100}}{p_{100}+p_{010}+p_{001}}$	$\frac{p_{001}+p_{010}}{p_{000}+p_{001}+p_{010}+p_{011}}$	$\frac{p_{100}}{p_{100}+p_{101}+p_{110}+p_{111}}$
2		$\frac{p_{110}+p_{101}}{p_{110}+p_{011}+p_{101}}$	$\frac{p_{011}}{p_{000}+p_{001}+p_{010}+p_{011}}$	$\frac{p_{101}+p_{110}}{p_{100}+p_{101}+p_{110}+p_{111}}$
3		1	0	$\frac{p_{111}}{p_{100}+p_{101}+p_{110}+p_{111}}$

Beliefs b	Accuracy Loss when $e_n = 0$	Accuracy Loss when $e_n = 1$
$b = P(e_n = 1\|s)$	$L(0,1) = \sum_b (0-b)^2 P[b\|e_n = 0, G = 1]$	$L(1,1) = \sum_b (1-b)^2 P[b\|e_n = 1, G = 1]$
0	$\left(\frac{p_{100}}{p_{100}+p_{010}+p_{001}}\right)^2 \times \frac{p_{001}+p_{010}}{p_{000}+p_{001}+p_{010}+p_{011}} + \left(\frac{p_{110}+p_{101}}{p_{110}+p_{011}+p_{101}}\right)^2 \times \frac{p_{011}}{p_{000}+p_{001}+p_{010}+p_{011}}$	$\left(\frac{p_{010}+p_{001}}{p_{100}+p_{010}+p_{001}}\right)^2 \times \frac{p_{100}}{p_{100}+p_{101}+p_{110}+p_{111}} + \left(\frac{p_{011}}{p_{110}+p_{011}+p_{101}}\right)^2 \times \frac{p_{101}+p_{110}}{p_{100}+p_{101}+p_{110}+p_{111}}$
$\frac{p_{100}}{p_{100}+p_{010}+p_{001}}$		
$\frac{p_{110}+p_{101}}{p_{110}+p_{011}+p_{101}}$		
1		

From the data in these tables we can compute expected accuracy loss. In the paper, our main theoretical measure of accuracy loss is Expected Accuracy Loss

$$E_{e_n}(L(e_n, G)) = L(0, G) \times P(e_n = 0) + L(1, G) \times P(e_n = 1)$$

which yields

$$\begin{aligned} E_{e_n}(L(e_n, 0)) = &\left(\frac{p_{100} + p_{110}}{p_{001} + p_{011} + p_{100} + p_{110}}\right)^2 \frac{p_{001} + p_{011}}{p_{000} + p_{001} + p_{010} + p_{011}}(p_{000} + p_{001} + p_{010} + p_{011}) \\ &+ \left(\frac{p_{001} + p_{011}}{p_{001} + p_{011} + p_{100} + p_{110}}\right)^2 \frac{p_{100} + p_{110}}{p_{100} + p_{101} + p_{110} + p_{111}}(p_{100} + p_{101} + p_{110} + p_{111}) \\ = &\frac{(p_{001} + p_{011})(p_{100} + p_{110})}{p_{001} + p_{011} + p_{100} + p_{110}} \end{aligned}$$

$$\begin{aligned} E_{e_n}(L(e_n, 1)) = &\left(\frac{p_{100}}{p_{100} + p_{010} + p_{001}}\right)^2 (p_{001} + p_{010}) + \left(\frac{p_{110} + p_{101}}{p_{110} + p_{011} + p_{101}}\right)^2 p_{011} \\ &+ \left(\frac{p_{010} + p_{001}}{p_{100} + p_{010} + p_{001}}\right)^2 p_{100} + \left(\frac{p_{011}}{p_{110} + p_{011} + p_{101}}\right)^2 (p_{101} + p_{110}) \\ = &\frac{(p_{010} + p_{001})p_{100}}{p_{100} + p_{010} + p_{001}} + \frac{p_{011}(p_{101} + p_{110})}{p_{110} + p_{011} + p_{101}} \end{aligned}$$

Having the loss functions, we must compute the covariance between expertise e_n and GAI boost e_g:

$$\begin{aligned} Cov(e_n, e_g) = &E[(e_n - \bar{e}_n)(e_g - \bar{e}_g)] \\ &= (0 - \bar{e}_n)(0 - \bar{e}_g) \times P(e_n = 0, e_g = 0) \\ &+ (1 - \bar{e}_n)(0 - \bar{e}_g) \times P(e_n = 1, e_g = 0) \\ &+ (0 - \bar{e}_n)(1 - \bar{e}_g) \times P(e_n = 0, e_g = 1) \\ &+ (1 - \bar{e}_n)(1 - \bar{e}_g) \times P(e_n = 1, e_g = 1) \end{aligned}$$

First note that

$$\begin{aligned} \bar{e}_n &= p_{100} + p_{101} + p_{110} + p_{111} \\ \bar{e}_g &= p_{010} + p_{110} + p_{011} + p_{111} \end{aligned}$$

And the joint probabilities are as follows:

Probabilities	Joint distribution
$P(e_n = 0, e_g = 0)$	$p_{000} + p_{001}$
$P(e_n = 1, e_g = 0)$	$p_{100} + p_{101}$
$P(e_n = 0, e_g = 1)$	$p_{010} + p_{011}$
$P(e_n = 1, e_g = 1)$	$p_{110} + p_{111}$

We are interested in the relationship between the covariance and accuracy loss, so we compute the Expected Accuracy Loss and the Covariance. Unlike the model in the paper, Expected Accuracy Loss does not depend directly on the covariance but on all the distributional parameters. We plot both on the (Expected Loss, Covariance) plane for each possible eight-tuple simplex of probabilities. Because there are an infinite amount, we pick distributions randomly from the eight-dimensional simplex. Figures A2 and A3 show the plot for a small number of distributions (about 1,000) and a large number of distributions (about 100,000), respectively. The figures corroborate that distributions with greater covariances typically entail lower accuracy loss when GAI is available than when GAI is not available. Due to the complexity of the problem, we use Python to draw the graph. The code is below.

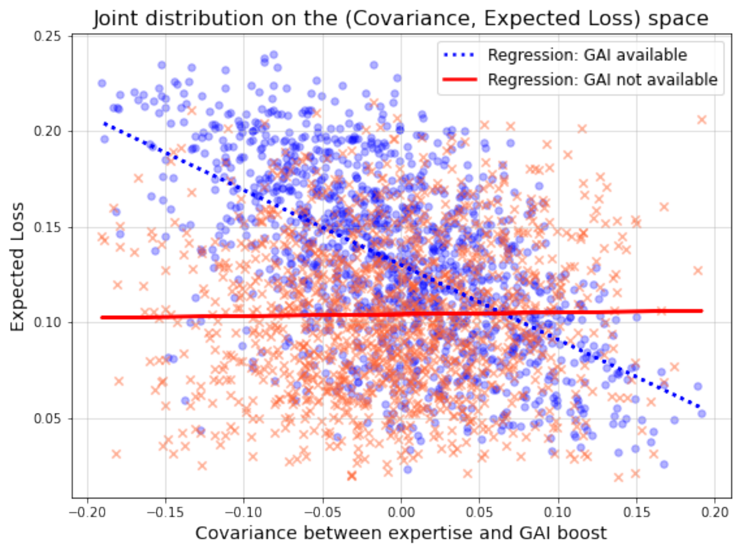


Figure A2: Accuracy Loss and Covariance, 1,319 randomly drawn distributions

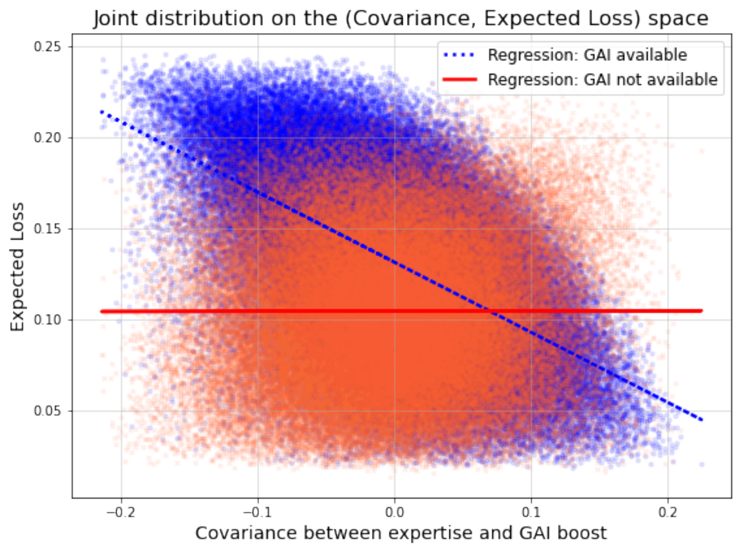


Figure A3: Accuracy Loss and Covariance, 128,401 randomly drawn distributions

A5 Python code for Accuracy Loss and Covariance Plots (Binary Model)

```
import numpy as np
import matplotlib.pyplot as plt
import time
from sklearn.linear_model import LinearRegression

start_time = time.time()

# Define the functions for the x-axis and y-axis
def x_function(p000, p001, p010, p011, p100, p101, p110, p111):
    return ((0-(p100+p101+p110+p111))*(0-(p010+p110+p011+p111))*(p000+p001) +
            (1-(p100+p101+p110+p111))*(0-(p010+p110+p011+p111))*(p100+p101) +
            (0-(p100+p101+p110+p111))*(1-(p010+p110+p011+p111))*(p010+p011) +
            (1-(p100+p101+p110+p111))*(1-(p010+p110+p011+p111))*(p110+p111))

# E_{e_n}[L(e_n,1)] expected loss when GAI available
def y_function(p000, p001, p010, p011, p100, p101, p110, p111):
    return (p100*(p010+p001))/(p100+p010+p001) + (p011*(p101+p110))/(p110+p011+p101)

# E_{e_n}[L(e_n,1)] expected loss when GAI not available
def z_function(p000, p001, p010, p011, p100, p101, p110, p111):
    return ((p001+p011)*(p100+p110))/(p001+p011+p100+p110)


# Number of random samples
# Different from number of distributions,
# because the distributions are a subset (the simplex))
#num_samples = 1000000000
num_samples = 10000000

# Initialize lists for x and y and z values
```

```
x_values = []
y_values = []
z_values = []


# Generate random samples for (p_{nge}) that satisfy constraints
for _ in range(num_samples):
    # Generate random values for p_{nge}
    p000, p001, p010, p011, p100, p101, p110 = np.random.uniform(0.01, 1, 7)  # Positive values
    if p000 + p001 + p010 + p011 + p100 + p101 + p110 < 1:  # Ensure p000 > 0
        p111 = 1 - (p000 + p001 + p010 + p011 + p100 + p101 + p110)  # p111 within the 8-simplex
        if p111 > 0:  # Ensure p111 is positive
            x = x_function(p000, p001, p010, p011, p100, p101, p110, p111)
            y = y_function(p000, p001, p010, p011, p100, p101, p110, p111)
            z = z_function(p000, p001, p010, p011, p100, p101, p110, p111)
            if y >= 0 and z >= 0:  # Ensure non-negative y-values abd z-values
                x_values.append(x)
                y_values.append(y)
                z_values.append(z)

# Convert lists to NumPy arrays when fitting the model
x_array = np.array(x_values).reshape(-1, 1)  # Reshape to 2D for sklearn
y_array = np.array(y_values)
z_array = np.array(z_values)

# Fit regression line for Dataset 1
model1 = LinearRegression()
model1.fit(x_array, y_array)
y_pred = model1.predict(x_array)

# Fit regression line for Dataset 2
model2 = LinearRegression()
model2.fit(x_array, z_array)
z_pred = model2.predict(x_array)

# Plot the parametric curve
# Create the plot
plt.figure(figsize=(8, 6))

# Plot the first dataset
plt.scatter(x_values, y_values, color='blue', s=30, alpha=0.3)
#s=40, alpha=0.5 best visualizes smaller num_samples.
#s=10, alpha=0.1 best visualizes large num_samples.

# Regression line for Dataset 1
plt.plot(x_values,
         y_pred,
         color='blue',
         linestyle=':',
         linewidth = 2.5,
         label='Regression: GAI available')

# Plot the second dataset
plt.scatter(x_values, z_values, c='#FF6433', s=40, alpha=0.5, marker='x')
```

```
#s=40, alpha=0.5 best visualizes smaller num_samples.
#s=10, alpha=0.1 best visualizes large num_samples.

# Regression line for Dataset 2
plt.plot(x_values,
         z_pred,
         color='red',
         linestyle='-',
         linewidth = 2.5,
         label='Regression: GAI not available')

# Add labels and title
plt.xlabel('Covariance between expertise and GAI boost', fontsize=14)
plt.ylabel('Expected Loss', fontsize=14)
plt.title('Joint distribution on the (Covariance, Expected Loss) space', fontsize=16)

# Add legend and grid
plt.legend(fontsize=12)
plt.grid(alpha=0.5)

# Show the plot
plt.tight_layout()
plt.savefig(f'Expected loss-covariance_{len(y_values)}.png', format="png")
plt.show()

end_time     = time.time()
elapsed_time = end_time - start_time

print(f"Elapsed time: {elapsed_time} seconds")
```

B Estimating σ_{ng} Covariance in our Experiment

The covariance between expertise and GAI boost (σ_{ng}) determines the direction of GAI effect on the outcome variables of interest. Here we develop a strategy for measuring the covariance σ_{ng} empirically.

Let $s^*(G) = a^*(G) + e_n + Ge_g + \epsilon$ be the level of the signal in equilibrium with GAI availability $G = 0, 1$. We begin with the following definition:

Definition 1. *Let $\Delta \equiv s^*(1) - s^*(0)$, the difference between the level of a subject's signal with GAI and without in equilibrium.*

The quantity Δ can be measured empirically for each sender by measuring pitches with and without access to GAI. The expertise of senders e_n is observed by the experimenters (but not receivers). Thus, we can derive the key covariance σ_{ng} by using Δ and e_n.

Proposition B1. *The covariance between Δ and expertise e_n equals σ_{ng}.*

Proof. Proposition B1. Note that the signal when GAI is available is $s^*(1) = a^*(1) + e_n + e_g + \epsilon$ and when GAI is not available is $s^*(0) = a^*(0) + e_n + \epsilon$, with $a^*(1)$ and $a^*(0)$ the effort in equilibrium, respectively. Recall that in equilibrium, the effort is constant; it does not depend on any of the random variables: $a^*(G) = r\pi(G)$, for GAI availability $G = 0, 1$.

The difference in the level of the signal with and without GAI is $\Delta = s^*(1) - s^*(0) = e_g + r(\pi(1) - \pi(0))$ (ϵ is independent of GAI availability). Because the second term $r(\pi(1) - \pi(0))$ does not depend on e_n or e_g, the covariance between Δ and e_n is simply σ_{ng}. $\square$

Using the previous result, we computed the covariance between Δ and expertise empirically. We created a dataset consisting of pairs of observations by the same sender, in the same domain, where the sender could use ChatGPT in one observation and could not use ChatGPT in the other. There were 16,089 such pairs. We then took the difference between the rating of the ChatGPT side of the pair and the non-ChatGPT side to find pairwise Δs. Finally, we took the covariance of the pairwise Δ variable and the sender's expertise in that domain and used bootstrapping to compute standard errors.

Our estimate of σ_{ng} is negative with a point estimate of -0.012, with a bootstrapped standard error of 0.0066. Our p-value for testing this estimate against zero is 0.072.

For non-English-speaking sender contexts, σ_{ng} was +0.056, with a bootstrapped standard error of .013, and a p-value below 0.001 when tested against zero.

For senders who *do* come from English-speaking contexts, σ_{ng} was -0.034, with a bootstrapped standard error of .0075, and a p-value below 0.001 when tested against zero.

C Experimental Design

Below we outline how we conducted our experimental design. Figure C4 summarizes the overall structure. We begin with tasking senders to write pitches with and without ChatGPT to convince senders in areas where they do have expertise and where they do not have expertise. This means that each sender writes four pitches. We do this across four domains in the entrepreneurship and hiring contexts: data science and management consulting in the hiring context and education and retail in the entrepreneurship context.

We then randomly show eight of the pitches one-by-one to each receiver in each of the four domains. Receivers evaluate the expertise, quality, and willingness to pay for costlier signals for each pitch. There is a 50 percent probability that senders could use ChatGPT for each of the eight pitches. The pitches that *could not* use ChatGPT are labeled on the screen. Figure C5 illustrates the possibilities given this randomization on the receiver end.

To interface with senders and receivers over Prolific, we developed apps using Adalo (`http://adalo.com`). The apps that we developed are essentially survey apps akin to Qualtrics, although they have a more sophisticated ability to interface with a database. In our case, we needed this database in our receiver app in order to host the senders' pitches and randomize them.

Figure C4: **Experiment layout**

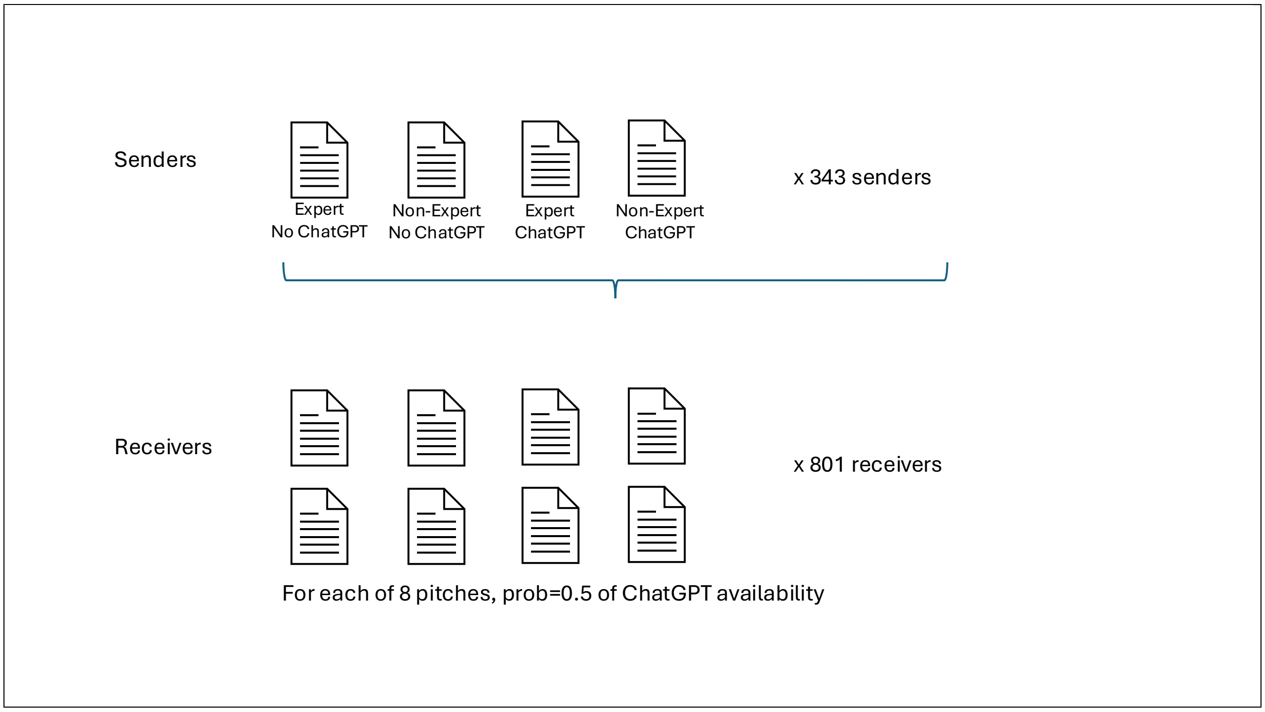
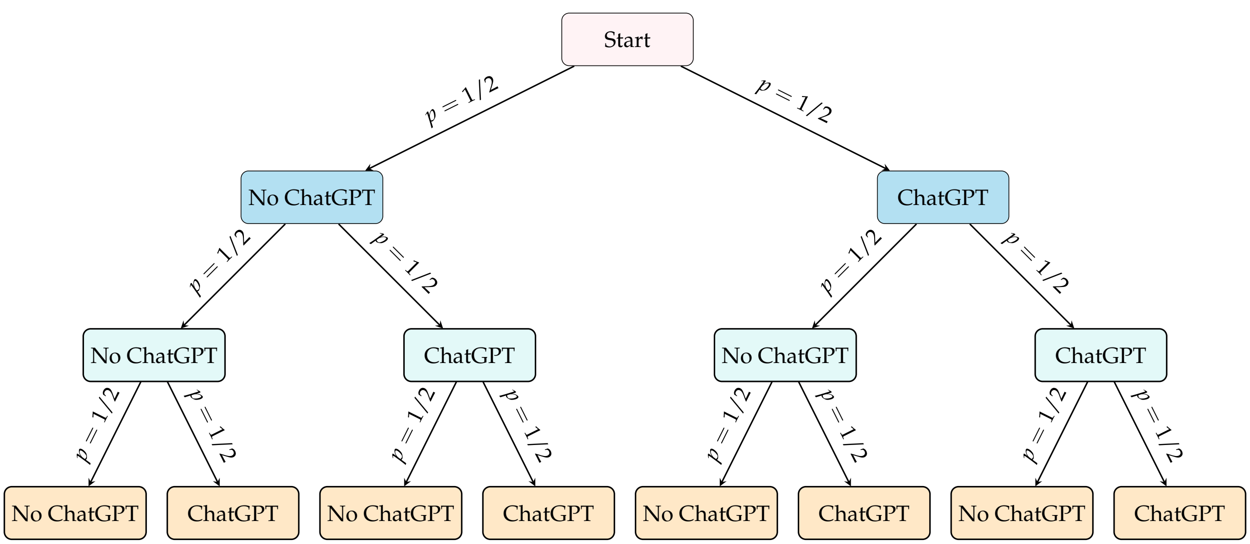


Figure C5: **Receiver Pitch Randomization: First 4 Assignments**



Notes: This figure shows how receivers are assigned their eight different senders' pitches to review. This diagram shows the first four assignments, and the remaining four continue the binary tree pattern above, with a $p = 1/2$ split between ChatGPT and No-ChatGPT after each assignment to a sender.

Below we show the specific instructions given to senders and receivers.

C1 Instructions for Senders

Below are the instructions that we gave to senders in the education industry in the entrepreneurship context. The instructions are similar for senders in the retail industry in the entrepreneurship context and in data science/management consulting backgrounds in the hiring context.

C1.1 Without ChatGPT Conditions

[Expert Task] For the education and training industry:

- Write a pitch to a potential investor (500-1500 characters) that describes a major problem that challenges the industry and a potential solution to address it.
- Explain your qualifications to tackle this problem and solution and make a persuasive case for them to invest in you.
- We will show your pitch to an actual investor in the industry, and your pitch will be judged on its quality.
- Please do not use ChatGPT to compose your pitch.

[Non-Expert Task] Now imagine that you would like to switch to the retail industry:

- Write a pitch (500-1500 characters) to a potential investor that describes a major problem that challenges the industry and a potential solution to address it.
- Explain your qualifications to tackle this problem and solution and make a persuasive case for them to invest in you.
- Please do not use ChatGPT to compose your pitch.

Of course, you are not currently in the retail industry. So you will need to make your actual experience sound as similar to retail as possible.

We will show your pitch to an actual investor in the industry, and your pitch will be judged on its quality.

C1.2 With ChatGPT Conditions

We would now like to encourage you to use ChatGPT, an AI-based chatbot, to enhance your two pitches. To do so:

1. Please go to https://openai.com/blog/chatgpt/.
2. Click the "try ChatGPT" button.
3. Sign up for a free account or log in with an existing account.
4. Type directions into the chat. For example, "write a startup pitch for the education and training industry."

In case you have not seen ChatGPT before, here are some suggestions:

1. Type "Write a startup pitch for the education and training industry." Then, edit or develop the output further to make sure it reflects your voice, qualifications, and message.
2. You can also copy your first pitch(es) into ChatGPT, and ask the program to improve the level of detail, specifics and/or vocabulary.

Feel free to use ChatGPT as you see fit to maximize your impression of the investors.

First, use ChatGPT to edit and improve your pitch for the education and training industry (repasted below).

Second, use ChatGPT to edit and improve your pitch for the retail industry (repasted below).

C2 Instructions for Receivers

Below we include instructions that we gave to investors in the education and training industry in the entrepreneurship context. The instructions are similar to those for receivers in the retail industry in the entrepreneurship context, as well as to those for receivers with data science/management consulting backgrounds in the hiring context.

C2.1 Background Information

In this exercise, you're going to read and evaluate eight startup pitches in the education and training sector. These letters were written by real entrepreneurs who are trying to get an investment.

Some of the letters were written by people who DO NOT have any relevant education and training experience. Their letters likely contain exaggerations of their qualifications and experiences—in an effort to convince you that they are worthy of getting investment.

Other letters were written by entrepreneurs who DO have an education and training background and whose qualifications are legitimate.

Both groups are being rewarded for convincing investors (such as yourselves) that they are qualified and worthy of getting funding—even though some of them are actually NOT. As such, they've tried their best to write a persuasive and convincing pitch.

We need your help finding good pitches and differentiating the two groups.

Before you begin, please note that some subjects have received help writing their pitches from AI (specifically ChatGPT). We are able to tell which pitches used the AI/ChatGPT assistance, and will provide you with this information about each pitch.

Please note that ChatGPT/AI was used both by qualified and unqualified entrepreneurs.

The AI assistance could possibly help qualified entrepreneurs convey their experience and expertise more clearly.

It could also help unqualified entrepreneurs exaggerate their experience.

In the pitches you will review, the qualified entrepreneurs were equally likely to use AI/ChatGPT assistance as the unqualified ones.

For each pitch, you will be asked to:

- Evaluate whether you believe the person was ever an entrepreneur in the education and training sector formerly (or currently).
- Suggest a decision about proceeding with the pitch for possible investment.
- Decide whether to conduct market research on the idea.
- After you submit your answers, we will verify which subjects actually have education and training entrepreneurship experience. Your answers to this survey will be judged based on this ground truth.

C2.2 Pitch Evaluation Questions

Please evaluate whether you believe the candidate was ever an entrepreneur in education and training formerly (or currently) on a probability of 0-100 (in buckets of 20, e.g., 0-20, 20-40...).

Please rate the following pitch on a scale of 1 to 5 (1=not convincing at all, likely would not get an interview for investment; 3=somewhat convincing, likely to be interviewed; 5=very convincing, likely to get interview and investment).

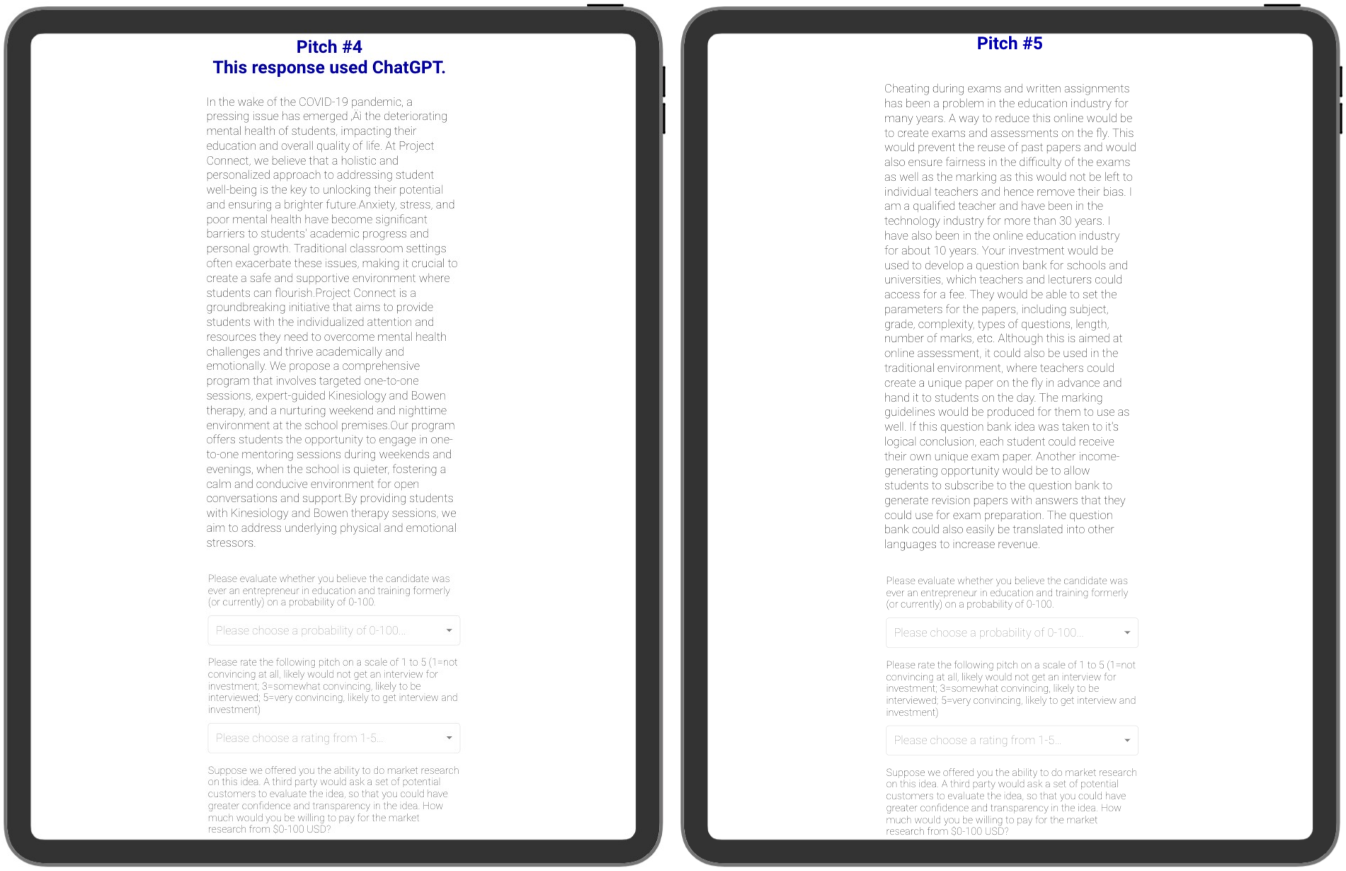


Figure C6

Suppose we offered you the ability to do market research on this idea. A third party would ask a set of potential customers to evaluate the idea, so that you could have greater confidence and transparency in the idea. How much would you be willing to pay for the market research from $0-100 USD? (in buckets of 20, for example, 0-20, 20-40...).

Figure C6 illustrates this evaluation task. The pitch on the left could use ChatGPT, while the one on the right could not. Each receiver evaluates eight of such screens.

D Additional Empirical Results

D1 Difference Between Hiring and Entrepreneurship Contexts

Table D1: **Summary Statistics by Setting**

	Startup	Hiring	Difference
Sender's True Expertise (0/1)	0.49	0.50	-.0074
Receiver's Belief Sender is an Expert (p)	0.43	0.47	-.033***
Variance of Receiver Belief, $p \cdot (1-p)$	0.17	0.17	.00034
Log(Variance of Receiver Belief)	-1.84	-1.84	.00091
Squared Error: [True Expertise - Receiver Belief]^2	0.30	0.31	-.01
Log(Squared Error)	-1.97	-1.92	-.049
Level of Signal (Rating 1-5 by Receiver)	2.78	2.96	-.19***
Log(Level of Signal, Rating)	0.90	0.98	-.082***
Level of Signal (NLP)	59.76	63.52	-3.8***
Log(NLP Rating)	4.03	4.12	-.089***
Variance of Rating	1.55	1.48	.07*
Log(Variance of Rating)	-0.75	-0.85	.098*
Variance of NLP Rating	248.48	187.63	61***
Log(Variance of NLP Rating)	4.05	3.82	.23***
Receiver WTP	38.60	38.07	.53
Receiver WTP (1-5 Rating)	2.43	2.40	.027
Log(Receiver WTP)	3.36	3.34	.014

Notes: This table contains summary statistics for our hiring and entrepreneurship settings. The average dependent variables are similar across the two settings.

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D2: **Effect Differences by Setting**

	(1) Error^2	(2) Variance of Receiver Belief	(3) Rating by Receiver	(4) NLP Rating	(5) Var of Rating	(6) Var of NLP
ChatGPT	-.0087	.003	.086*	12***	-.12*	-165***
	(.01)	(.0028)	(.045)	(.36)	(.063)	(16)
ChatGPT× Startup	.019	.0058	.06	9.2***	-.019	-42
	(.015)	(.0037)	(.063)	(.59)	(.085)	(31)
FEs	All	All	All	All	All	All
R^2	.2	.26	.37	.72	.27	.51
DV Mean	.31	.17	2.9	62	1.5	218
Observations	6,438	6,438	6,438	6,440	6,438	6,440

Notes: This table shows differences in the main dependent variables by whether the task is in the entrepreneurship (Startup=1) or hiring (Startup=0) settings. The columns include sender, receiver, domain, and order fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D2 Contingency Tables

Table D3: **Absolute Error Contingency Table**

	Abs(Error) = |Predicted - Actual Expertise|					
	.1	.3	.5	.7	.9	Total
No ChatGPT	703	679	643	599	546	3,170
ChatGPT	617	735	715	686	515	3,268
Total	1,320	1,414	1,358	1,285	1,061	6,438

Pearson $\chi^2(4) = 16.9464$, $p = 0.002$

Notes: This is a contingency table expressing raw counts of observations in all categories. We examine ChatGPT× squared error. We calculate predicted expertise by using the midpoint of all five probability bins [0-.2], [0.2-.4], [.4-.6], [.6-.8] and [.8-1]. The null hypothesis of the χ^2 test is: The distribution of the error categories is the same whether ChatGPT=0 or ChatGPT=1 (in other words, ChatGPT is statistically independent of error).

Table D4: **Accuracy Contingency Table (Raw Error)**

	Error = Predicted - Actual Expertise										
	-.9	-.7	-.5	-.3	-.1	.1	.3	.5	.7	.9	Total
No ChatGPT	364	285	334	377	210	493	302	309	314	182	3,170
ChatGPT	357	312	373	389	184	433	346	342	374	158	3,268
Total	721	597	707	766	394	926	648	651	688	340	6,438

Pearson $\chi^2(9) = 19.3317$, $p = 0.023$

Notes: This is a contingency table expressing raw counts of observations in all categories. We examine ChatGPT$\times$ squared error. We calculate predicted expertise by using the midpoint of all five probability bins [0-.2], [0.2-.4], [.4-.6], [.6-.8] and [.8-1]. The null hypothesis of the χ^2 test is: The distribution of the error categories is the same whether ChatGPT=0 or ChatGPT=1 (in other words, ChatGPT is statistically independent of error).

Table D5: **Confusion Matrix (ChatGPT Senders)**

	Predicted Expertise					
	$[0, 0.2]$	$(0.2, 0.4]$	$(0.4, 0.6]$	$(0.6, 0.8]$	$(0.8, 1]$	Total
Actual Expert	357	312	373	389	184	$1,615$
Actual Non-Expert	433	346	342	374	158	$1,653$
Total	790	658	715	763	342	$3,268$

Notes: This figure shows a confusion matrix for all predictions made by receivers/screeners in the data on senders who *could* use ChatGPT in their pitches.

Table D6: **Confusion Matrix (Non-ChatGPT Senders)**

	Predicted Expertise					
	$[0, 0.2]$	$(0.2, 0.4]$	$(0.4, 0.6]$	$(0.6, 0.8]$	$(0.8, 1]$	Total
Actual Expert	364	285	334	377	210	$1,570$
Actual Non-Expert	493	302	309	314	182	$1,600$
Total	857	587	643	691	392	$3,170$

Notes: This figure shows a confusion matrix for all predictions made by receivers/screeners in the data on senders who could *not* use ChatGPT in their pitches.

D3 ChatGPT and Beliefs of Expertise

Table D7: **ChatGPT and Beliefs of Expertise**

	(1) Pr Sender is Expert	(2) Pr Sender is Expert
ChatGPT	.0037	.0076
	(.0075)	(.0073)
FEs	-	All
R^2	.000047	.41
DV Mean	.45	.45
Observations	6,438	6,438

Notes: The table shows results from regressions predicting the probability that receivers expect senders to be experts from their access to ChatGPT. Robust standard errors (in parentheses) are clustered at the receiver level.+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D4 Predictors of Highly-Rated Pitches

What factors prompt evaluators to give higher signal levels and beliefs of expertise to pitches? To assess this, we conduct topic modeling on the sender pitches in each context using an LDA approach. The LDA model detects clusters of words in the corpus of text that constitute different topics (Pritchard et al., 2000; Blei et al., 2003). We then use ChatGPT to help us label these topics based on the cluster of words. This approach allows us to categorize the pitches within each context into categories of topics. We can then measure which topics are more predictive of the signal levels and beliefs of expertise that evaluators ascribe.

Table D8 shows that in the consulting recruitment context, mentioning "strategic consulting" and "business analysis" topics are most predictive of higher signal levels and beliefs of expertise, while mentioning "data science" topics predicts lower levels. Table D9 shows that in the data science recruitment context, mentioning topics related to "teamwork and management skills" predicts lower ratings and beliefs of expertise. Table D10 reveals that in the retail entrepreneurship context, "training and customer experience" topics predict higher signal levels and beliefs of expertise, while "product-focused solutions" topics predict lower ones. Lastly, Table D11 shows that mentioning "customer experience" topics predicts higher beliefs of expertise in the education entrepreneurship context.

Table D8: **Topics that predict higher rating in the consulting recruitment context**

	(1) Rating	(2) Rating	(3) Rating	(4) Rating	(5) Rating	(6) Belief Expert	(7) Belief Expert	(8) Belief Expert	(9) Belief Expert	(10) Belief Expert
General Business Experience	-0.104*					-0.013				
	(0.052)					(0.012)				
Strategic Consulting/Client Management		0.199**					0.048***			
		(0.064)					(0.014)			
Project Roles and Research			0.047					0.004		
			(0.098)					(0.020)		
Data Science				-0.232***					-0.056***	
				(0.059)					(0.014)	
Business Analysis					0.139**					0.033*
					(0.053)					(0.013)
_cons	3.010***	2.918***	2.962***	3.022***	2.885***	0.477***	0.460***	0.471***	0.485***	0.452***
	(0.022)	(0.015)	(0.008)	(0.014)	(0.031)	(0.005)	(0.003)	(0.002)	(0.003)	(0.007)
N	2372	2372	2372	2372	2372	2372	2372	2372	2372	2372
Evaluator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full	Full	Full	Full	Full	Full

Notes: This table shows topics that predict receiver ratings using an LDA topic modeling approach. Robust standard errors (in parentheses) are clustered at the evaluator level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D9: **Topics that predict higher ratings in the data science recruitment context**

	(1) Rating	(2) Rating	(3) Rating	(4) Rating	(5) Rating	(6) Belief Expert	(7) Belief Expert	(8) Belief Expert	(9) Belief Expert	(10) Belief Expert
Data Science/Analytical	0.105					0.0186				
	(1.49)					(1.18)				
Client-Oriented Roles		0.0741					0.0224			
		(1.11)					(1.47)			
Technical Management			-0.103					-0.0459*		
			(-1.29)					(-2.52)		
Science Teamwork				-0.409***					-0.0797***	
				(-4.74)					(-3.86)	
Strategic Projects					0.0149					0.0187
					(0.13)					(0.74)
_cons	2.934***	2.938***	2.999***	3.040***	2.976***	0.462***	0.458***	0.479***	0.482***	0.468***
	(101.23)	(82.73)	(174.48)	(227.46)	(302.88)	(71.64)	(56.74)	(122.33)	(150.76)	(215.15)
N	1563	1563	1563	1563	1563	1563	1563	1563	1563	1563
Evaluator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full	Full	Full	Full	Full	Full

Notes: This table shows topics that predict receiver ratings using an LDA topic modeling approach. Robust standard errors (in parentheses) are clustered at the evaluator level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D10: **Topics that predict higher ratings in the retail startup context**

	(1) Rating	(2) Rating	(3) Rating	(4) Rating	(5) Rating	(6) Belief Expert	(7) Belief Expert	(8) Belief Expert	(9) Belief Expert	(10) Belief Expert
Retail Training	0.117^{+}					0.0288^{+}				
	(1.69)					(1.93)				
Retail/Customer Exp.		0.0971					0.0311^{*}			
		(1.47)					(2.11)			
Fashion/Product-Focused			-0.262^{***}					-0.0677^{***}		
			(-3.66)					(-4.60)		
Education/Workforce				-0.0588					-0.0238	
				(-0.73)					(-1.27)	
Retail Challenges					0.00682					0.00395
					(0.10)					(0.27)
_cons	2.747^{***}	2.750^{***}	2.853^{***}	2.792^{***}	2.781^{***}	0.427^{***}	0.426^{***}	0.454^{***}	0.440^{***}	0.435^{***}
	(130.38)	(125.20)	(148.34)	(208.63)	(153.08)	(94.38)	(86.86)	(114.72)	(142.05)	(111.43)
N	1637	1637	1637	1637	1637	1637	1637	1637	1637	1637
Evaluator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full	Full	Full	Full	Full	Full

Notes: This table shows topics that predict receiver ratings using an LDA topic modeling approach. Robust standard errors (in parentheses) are clustered at the evaluator level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D11: **Topics that predict higher ratings in the education startup context**

	(1) Rating	(2) Rating	(3) Rating	(4) Rating	(5) Rating	(6) Belief Expert	(7) Belief Expert	(8) Belief Expert	(9) Belief Expert	(10) Belief Expert
Small Business	0.0871					0.0325^{*}				
	(1.32)					(2.21)				
Customer Experience		0.162^{**}					0.0410^{***}			
		(3.00)					(3.53)			
Business Challenges			-0.0989^{+}					-0.0234^{*}		
			(-1.88)					(-2.18)		
Education/Skills				0.0199					-0.000106	
				(0.33)					(-0.01)	
Specialized Training					-0.0337					-0.00328
					(-0.50)					(-0.24)
_cons	2.731^{***}	2.694^{***}	2.793^{***}	2.746^{***}	2.761^{***}	0.425^{***}	0.419^{***}	0.443^{***}	0.434^{***}	0.435^{***}
	(159.27)	(134.00)	(134.02)	(113.66)	(186.80)	(111.29)	(96.89)	(103.99)	(84.47)	(143.83)
N	2363	2363	2363	2363	2363	2363	2363	2363	2363	2363
Evaluator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full	Full	Full	Full	Full	Full

Notes: This table shows topics that predict receiver ratings using an LDA topic modeling approach. Robust standard errors (in parentheses) are clustered at the evaluator level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D5 Effects of ChatGPT on Pitch Topics

How does ChatGPT impact the content of pitches? If ChatGPT merely prompted senders to revise their pitches, then we would expect there to be no differences in the substantive content, but rather simply in the writing. To test this, we assess how pitches created with or without ChatGPT vary in the topics they mention using the topic categories created in the LDA exercise in Section D4. Tables D12– D15 show that the topics that pitches mention meaningfully vary between those that can use ChatGPT versus those that cannot. This is particularly the case in the consulting recruitment and education startup contexts, which suggests that the ChatGPT effects reflect more than revisions.

Table D12: **ChatGPT's effect on topics in the consulting recruitment context**

	(1) General Business Experience	(2) Strategic Consulting Client Management	(3) Project Role Research	(4) Data Science	(5) Business Analysis
ChatGPT	-0.007	-0.035^{+}	0.009	-0.045*	-0.088***
	(0.022)	(0.018)	(0.013)	(0.018)	(0.022)
Expert	-0.051*	0.128***	-0.025*	0.020	-0.056**
	(0.020)	(0.019)	(0.012)	(0.017)	(0.020)
_cons	0.454***	0.187***	0.090***	0.246***	0.642***
	(0.012)	(0.011)	(0.007)	(0.010)	(0.012)
N	2372	2372	2372	2372	2372
Evaluator FE	Yes	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full

Notes: This table shows the relationship between ChatGPT and topics mentioned in pitches using an LDA topic modeling approach. Robust standard errors (in parentheses) are clustered at the evaluator level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D13: **ChatGPT's effect on topics in the data scientist recruitment context**

	(1) Data Science Analytical	(2) Client-Oriented Roles	(3) Technical Management	(4) Science Teamwork	(5) Strategic Projects
ChatGPT	0.015	-0.005	-0.002	0.004	0.012
	(0.026)	(0.029)	(0.021)	(0.019)	(0.016)
Expert	-0.135***	0.034	-0.147***	-0.002	0.014
	(0.025)	(0.028)	(0.022)	(0.021)	(0.015)
_cons	0.469***	0.516***	0.289***	0.154***	0.073***
	(0.017)	(0.019)	(0.015)	(0.015)	(0.011)
N	1563	1563	1563	1563	1563
Evaluator FE	Yes	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full

Notes: This table shows the relationship between ChatGPT and topics mentioned in pitches using an LDA topic modeling approach. Robust standard errors (in parentheses) are clustered at the evaluator level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D14: **ChatGPT's effect on topics in the retail startup context**

	(1) Retail Training	(2) Retail Customer Exp.	(3) Fashion Product-Focused	(4) Education Workforce	(5) Retail Challenges
ChatGPT	-0.009	0.062*	0.016	-0.026	-0.020
	(0.024)	(0.026)	(0.024)	(0.020)	(0.023)
Expert	-0.031	-0.047+	-0.069**	-0.046**	0.133***
	(0.023)	(0.026)	(0.023)	(0.017)	(0.021)
_cons	0.325***	0.325***	0.298***	0.204***	0.202***
	(0.018)	(0.020)	(0.019)	(0.014)	(0.018)
N	1637	1637	1637	1637	1637
Evaluator FE	Yes	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full

Notes: This table shows the relationship between ChatGPT and topics mentioned in pitches using an LDA topic modeling approach. Robust standard errors (in parentheses) are clustered at the evaluator level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D15: **ChatGPT's effect on topics in the education startup context**

	(1) Small Business	(2) Customer Experience	(3) Business Challenges	(4) Education Skill	(5) Specialized Training
ChatGPT	0.025	0.033	0.000	-0.115***	-0.059**
	(0.019)	(0.021)	(0.021)	(0.021)	(0.018)
Expert	0.047*	-0.013	-0.016	-0.051*	0.027
	(0.019)	(0.021)	(0.021)	(0.021)	(0.018)
_cons	0.226***	0.367***	0.404***	0.471***	0.227***
	(0.011)	(0.013)	(0.012)	(0.013)	(0.011)
N	2363	2363	2363	2363	2363
Evaluator FE	Yes	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full	Full

Notes: This table shows the relationship between ChatGPT and topics mentioned in pitches using an LDA topic modeling approach. Robust standard errors (in parentheses) are clustered at the evaluator level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D6 What costlier signals do people prefer?

Our experimental results show that, on average, ChatGPT induces receivers to demand costlier signals than text. In our experimental design, we described such costlier signals as market research in the entrepreneurship context and background investigations in the recruitment context. Are these the top ones people demand? Are there others?

To investigate this, we send out a survey to our original Prolific sample, as well as a broader population. Tables D16– D17 show that, among those who indicate they have experience in investing, market research on the business idea is indeed the most commonly chosen costlier signal. This is the case whether looking at our original survey population or a broader sample and in English-speaking or non-English-speaking contexts. The next most commonly chosen costlier signals are the entrepreneur's background and the company's prior financial/user data.

Among those with hiring experience, we find that an interview is the most common costlier signal in both the original and broader survey populations and in English-speaking and non-English-speaking contexts. The next most commonly chosen signals are the candidate's background and job assessment.

These trends suggest that there is a convergence in preferences for costlier signals across populations. They also indicate that both an entrepreneur's and job candidate's background, while not the most-commonly preferred costlier signal, is among the top of choices. This leaves the possibility that investors and hiring managers might pay more attention to individual observables—like race, gender, and geography—that may be seen to correlate with expertise, posing a risk for discriminatory outcomes. We hope that future work explores these possibilities.

Table D16: **Costlier Signals from Initial vs. Broader Prolific Sample**

	(1)				
	Broader Obs	Broader Mean	Initial Obs	Initial Mean	Difference
Entrep Background	281	0.24	290	0.16	0.08*
Entrep Data	281	0.19	290	0.22	-0.04
Entrep Interview	281	0.08	290	0.16	-0.08**
Entrep Market Research	281	0.43	290	0.41	0.02
Entrep References	281	0.05	290	0.04	0.00
Hiring Background Investigation	354	0.09	335	0.07	0.02
Hiring Interview	354	0.42	335	0.51	-0.08*
Hiring Job Assessment	354	0.18	335	0.20	-0.02
Hiring Candidate Background	354	0.21	335	0.14	0.07*
Hiring References	354	0.09	335	0.08	0.02
N	871				

Notes: This table shows summary statistics for the initial vs. broader Prolific sample with investing or hiring experience. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D17: **Costlier Signals from English- vs. Non-English-Speaking Countries**

	(1)				
	Non-English Obs	Non-English Mean	English Obs	English Mean	Difference
Entrep Background	101	0.20	470	0.20	-0.00
Entrep Data	101	0.22	470	0.20	0.01
Entrep Interview	101	0.09	470	0.13	-0.04
Entrep Market Research	101	0.48	470	0.41	0.07
Entrep References	101	0.02	470	0.05	-0.03
Hiring Background Investigation	133	0.07	556	0.08	-0.02
Hiring Interview	133	0.51	556	0.45	0.06
Hiring Job Assessment	133	0.17	556	0.20	-0.02
Hiring Candidate Background	133	0.15	556	0.18	-0.03
Hiring References	133	0.10	556	0.08	0.02
N	871				

Notes: This table shows summary statistics for the Prolific sample from non-English-speaking vs. English-speaking countries with investing or hiring experience.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D7 Experts in non-English-speaking contexts are more likely to update their pitch with ChatGPT.

Figure 2 shows that the covariance term for senders from non-English-speaking contexts is more positive than that of senders from English-speaking contexts. This implies that ChatGPT is helping experts more than non-experts in non-English-speaking contexts than in others. This would also imply that there would be bigger updates in pitches with ChatGPT between experts vs. non-experts in non-English-speaking vs. English-speaking contexts.

To test this, we assess the extent to which ChatGPT results in a substantive change in the pitches using an LLM and compare this change for experts vs. non-experts in English-speaking vs. non-English-speaking sender contexts. Specifically, we prompt an LLM (using the OpenAI API) to evaluate the non-ChatGPT and ChatGPT pitches for each sender: "Are the responses focused on different subjects? Respond with 'Yes' or 'No.'" This gives us a binary variable that is our main dependent variable.

We then assess how this binary dependent variable varies by expertise in English-speaking vs. non-English-speaking sender contexts. On average, 9 percent of the pitches exhibit a subject change with ChatGPT.

Table D18 reports how this varies by expertise. It shows that, on average, there is no difference in the rate of changing topics with ChatGPT between experts and non-experts (columns 1–2). Consistent with the idea that ChatGPT drives larger differences in updating between experts and non-experts in non-English-speaking contexts where the covariance is higher, experts are more likely to update in non-English-speaking sender contexts (column 3) than in English-speaking contexts (column 4). And the difference is statistically meaningful (column 5).

These results suggest that ChatGPT *does* in fact change the substance of the pitches for a non-trivial share of senders. And it does so more for experts in non-English-speaking contexts, which may help account for the positive covariance that we observe in this population.

Table D18: **ChatGPT-induced pitch changes and expertise in non-English-speaking contexts**

	(1) Change	(2) Change	(3) Change	(4) Change	(5) Change
Expert	-0.003	-0.003	0.043*	-0.020	-0.020
	(0.016)	(0.016)	(0.021)	(0.021)	(0.021)
Expert x Non-English					0.063*
					(0.030)
_cons	0.093***	0.093***	0.011	0.124***	0.093***
	(0.016)	(0.008)	(0.011)	(0.010)	(0.008)
N	686	686	186	500	686
Sender FE	No	Yes	Yes	Yes	Yes
Setting FE	No	Yes	Yes	Yes	Yes
Sample	Full	Full	Non-English	English	Full

Notes: This table shows how subject changes in pitches induced by ChatGPT vary by expertise and whether senders are based in non-English-speaking contexts.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D8 Percentile/Ranking-Based Outcomes

Table D19: **Effects of ChatGPT & Expertise on Ratings Percentile**

	(1) Signal Level Percentile	(2) Signal Level Percentile	(3) Signal Level Percentile	(4) Signal Level Percentile
ChatGPT	.019**	.023***	.019**	.023***
	(.0078)	(.0074)	(.0078)	(.0074)
Experience			.028***	.035***
			(.007)	(.0073)
FEs	-	All	-	All
R^2	.0088	.35	.011	.35
DV Mean	.5	.5	.5	.5
Observations	6,440	6,440	6,440	6,440

Notes: The table shows results from regressions predicting the percentile of ratings using ChatGPT and expertise. The columns using all fixed effects are outlined in Equation 3 and include sender, receiver, domain, and order fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table D20: **Effects of ChatGPT & Expertise on P(Expertise) Percentile**

	(1) P(Expert) Percentile	(2) P(Expert) Percentile	(3) P(Expert) Percentile	(4) P(Expert) Percentile
ChatGPT	.0057	.01	.0057	.01
	(.008)	(.0076)	(.008)	(.0076)
Experience			.04***	.048***
			(.0069)	(.0072)
FEs	-	All	-	All
R^2	.0038	.33	.0086	.34
DV Mean	.5	.5	.5	.5
Observations	6,440	6,440	6,440	6,440

Notes: The table shows results from regressing the percentile of predicted expertise (p) on ChatGPT and expertise. The columns using all fixed effects are outlined in Equation 3 and include sender, receiver, domain, and order fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D9 Results using Between-Receiver Variation Only

In our main results, each receiver examines both pitches with and without access to GAI. In this sense, we have a "within-receiver" design that uses variation in GAI availability within the same receiver over multiple evaluations. This permits fixed effects for receivers.

In this section, we present a "between-receiver" design. We examine a subset of our data that includes i) each receiver's initial evaluation, and ii) any subsequent evaluations that similarly can use (or not use) GAI before the receiver is given an opposite-GAI pitch to evaluate. Given this, all variation in ChatGPT is between receivers, and we must remove receiver fixed effects. Below we reproduce our main tables using only the variation in ChatGPT from this subsample.

Table D21: **Between Subjects Only: Summary Statistics by ChatGPT Availability**

	With ChatGPT	No ChatGPT	Difference
Sender's True Expertise (0/1)	0.47	0.51	-.036
Receiver's Belief Sender is an Expert (p)	0.45	0.43	.02
Variance of Receiver Belief, $p \cdot (1-p)$	0.18	0.17	.011***
Log(Variance of Receiver Belief)	-1.81	-1.89	.073***
Squared Error: [True Expertise - Receiver Belief]^2	0.30	0.32	-.016
Log(Squared Error)	-1.91	-1.96	.053
Level of Signal (Rating 1-5 by Receiver)	2.84	2.73	.11*
Log(Level of Signal, Rating)	0.94	0.88	.064**
Level of Signal (NLP)	69.27	52.68	17***
Log(NLP Rating)	4.22	3.91	.31***
Variance of Rating	1.35	1.57	-.22***
Log(Variance of Rating)	-1.15	-0.57	-.58***
Variance of NLP Rating	125.61	287.81	-162***
Log(Variance of NLP Rating)	3.51	4.51	-1***
Receiver WTP	38.03	36.15	1.9
Receiver WTP (1-5 Rating)	2.40	2.31	.094
Log(Receiver WTP)	3.35	3.26	.089**

Notes: This table contains summary statistics for ChatGPT and non-ChatGPT messages. $N = 1,636$ total sender×receiver pairs. The stars are given by the p-values of the t-test of a difference in means with zero. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D22: **Between Subjects Only: ChatGPT Increases Screening Errors**

	(1) Variance of Receiver Belief $p \cdot (1-p)$	(2) Variance of Receiver Belief $p \cdot (1-p)$	(3) Log Variance of Receiver Belief	(4) Log Variance of Receiver Belief	(5) Error^2	(6) Error^2	(7) Log(Error^2)	(8) Log(Error^2)
ChatGPT	.011***	.011**	.073***	.076**	-.016	.0029	.053	.15
	(.0037)	(.0049)	(.025)	(.032)	(.014)	(.019)	(.08)	(.11)
FEs	-	All	-	All	-	All	-	All
R^2	.0065	.25	.0067	.25	.00079	.27	.00029	.27
DV Mean	.17	.17	-1.8	-1.8	.31	.31	-1.9	-1.9
Observations	1,636	1,636	1,636	1,636	1,636	1,636	1,636	1,636

Notes: The table shows results from regressing measures of screening errors on ChatGPT. The columns using all fixed effects are outlined in Equation 3 and include sender and domain fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D23: **Between Subjects Only: ChatGPT Increases the Level of the Signal**

	(1) Rating by Receiver	(2) Rating by Receiver	(3) Log Rating by Receiver	(4) Log Rating by Receiver	(5) NLP Rating	(6) NLP Rating	(7) Log NLP	(8) Log NLP
ChatGPT	.11	-.041	.064**	.0075	17***	17***	.31***	.31***
	(.072)	(.09)	(.03)	(.038)	(.75)	(.8)	(.015)	(.015)
FEs	-	Full	-	Full	-	Full	-	Full
R^2	.0021	.29	.0039	.28	.26	.81	.26	.81
DV Mean	2.8	2.8	.91	.91	62	62	4.1	4.1
Observations	1,636	1,636	1,636	1,636	1,638	1,638	1,638	1,638

Notes: The table shows results from regressing the level of the signal (as measured by evaluators' ratings and NLP) on ChatGPT. The columns using all fixed effects are outlined in Equation 3 and include sender and domain fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D24: **Between Subjects Only: ChatGPT Lowers Dispersion of Signals**

	(1) Var. of Rating	(2) Var. of Rating	(3) Log Var. of Rating	(4) Log Var. of Rating	(5) Var. of NLP	(6) Var. of NLP	(7) Log Var. of NLP	(8) Log Var. of NLP
ChatGPT	-.22**	-.3**	-.58***	-.71***	-161***	-121***	-.99***	-.9***
	(.09)	(.12)	(.11)	(.16)	(31)	(36)	(.098)	(.14)
FEs	-	All	-	All	-	All	-	All
R^2	.0055	.21	.016	.21	.02	.8	.053	.65
DV Mean	1.4	1.4	-.89	-.89	198	198	4	4
Observations	1,634	1,634	1,634	1,634	1,636	1,636	1,636	1,636

Notes: The table shows results from regressing the variance of the signal (as measured by evaluators' ratings and NLP) on ChatGPT. The columns using all fixed effects are outlined in Equation 3 and include sender and domain fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D25: **Between Subjects Only: ChatGPT Raises Demand for Costlier Signals**

	(1) WTP	(2) WTP	(3) Log WTP	(4) Log WTP	(5) WTP, 1-5 Scale	(6) WTP, 1-5 Scale	(7) Log WTP, 1-5 Scale	(8) Log WTP, 1-5 Scale
ChatGPT	1.9	.58	.089	.059	.094	.029	.058	.036
	(1.9)	(2.4)	(.055)	(.069)	(.093)	(.12)	(.039)	(.05)
FEs	-	All	-	All	-	All	-	All
R^2	.0013	.25	.0028	.25	.0013	.25	.0024	.25
DV Mean	37	37	3.3	3.3	2.4	2.4	.69	.69
Observations	1,636	1,636	1,636	1,636	1,636	1,636	1,636	1,636

Notes: The table shows results from regressing receivers' willingness to pay for costlier signals on ChatGPT. The columns using all fixed effects are outlined in Equation 3 and include sender and domain fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D26: **Between Subjects Only: ChatGPT and Non-English-Speaking Sender Contexts**

	(1)	(2)	(3)	(4)	(5)	(6)
	Log(Error^2)	Log Variance of Receiver Belief	Log Rating by Receiver	Log NLP	Log Var. of Rating	Log Var. of NLP
ChatGPT	.27**	.094**	-.011	.34***	-.65***	-.69***
	(.13)	(.037)	(.042)	(.018)	(.19)	(.18)
ChatGPT× Non-English	-.4	-.06	.059	-.12***	-.22	-.68**
	(.25)	(.068)	(.079)	(.034)	(.34)	(.28)
FEs	All	All	All	All	All	All
R^2	.27	.25	.28	.81	.21	.66
DV Mean	-1.9	-1.8	.91	4.1	-.89	4
Observations	1,636	1,636	1,636	1,638	1,636	1,638

Notes: The table shows results from regressing the main dependent variables on ChatGPT and the interaction between ChatGPT and whether the sender is from a non-English-speaking country. The columns using all fixed effects are outlined in Equation 3 and include sender and domain fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D10 Sample Survey

Do our senders and receivers have sufficient experience to complete the experimental exercises? While we cannot directly validate the skills of our subjects, we can compare their self-reported credentials to those of the broader Prolific pool. To do so, we send a follow-up survey to our initial 1,116 Prolific respondents who reported a credible Prolific ID and were still registered on the survey platform. We also sent the survey to the same number of prolific respondents from the broader pool. The survey elicited factors correlated with expertise, such as income, education, and years of work experience in the focal areas of the experiment. We also asked subjects about their experience with ChatGPT. We only kept fully completed survey responses. Over half of our initial subjects fully completed the survey in the final results.

Table D27 reports the results. It shows that our initial pool has a higher income, is more likely to have a graduate degree, and has more years of total work experience and specific experience in the relevant focus areas of the experiment. This suggests that respondents in the experiment had more expertise than the average Prolific user and are therefore more likely to possess the necessary knowledge for the experiment. Interestingly, our initial pool reports using ChatGPT marginally less intensively than the broader pool, suggesting that it was less likely to use ChatGPT even in non-ChatGPT scenarios in our experiment.

Tables D28-D29 further show that senders and receivers have more experience in the domains of their respective experimental contexts than do others. This holds whether looking at the original sample or the full sample pool. Table D28 shows that in the entrepreneurship context, senders have more years of experience as entrepreneurs than others (column 1–2). Receivers in the entrepreneurship context generally have more years of experience in investing (column 2–3). Senders and receivers in the education entrepreneurship context have more years of experience in the education sector (columns 5–6). Those in the retail entrepreneurship context have more years of experience in retail (columns 7–8).

Table D29 shows that the same trends hold for recruitment. Receivers in the recruitment context have more years of experience in human resources (column 1–2). Senders and receivers in the data science recruitment context have more years of experience in data science (column 3–4). And those in the consulting recruitment context have more years of experience in consulting (columns 5–6).

Together, these results suggest that our subjects, at least relative to the original and broader Prolific population, have more experience in the domains core to their experimental context.

Table D27: **Initial vs. Broader Prolific Sample**

	(1)				
	Broader Obs	Broader Mean	Initial Obs	Initial Mean	Difference
Income USD	1053	41049.38	624	56169.87	-1.5e+04***
Grad Degree	1053	0.21	624	0.34	-0.12***
Work Exp	1053	11.02	624	14.67	-3.65***
Entrep Exp	1053	1.75	624	3.13	-1.38***
Investing Exp	1053	0.82	624	2.16	-1.33***
Education Exp	1053	3.35	624	4.67	-1.32***
Retail Exp	1053	2.21	624	3.87	-1.66***
HR Exp	1053	1.28	624	3.12	-1.85***
Data Science Exp	1053	1.12	624	2.37	-1.25***
Consulting Exp	1053	1.24	624	2.35	-1.11***
ChatGPT Use	1053	1.52	624	1.27	0.24**
N	1677				

Notes: This table shows summary statistics for the initial vs. broader Prolific sample.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D28: **Domain experience in the entrepreneurship context**

	(1) Entrep. Years	(2) Entrep. Years	(3) Investing Years	(4) Investing Years	(5) Education Years	(6) Education Years	(7) Retail Years	(8) Retail Years
Entrepreneur	2.622***	2.622***						
	(0.637)	(0.637)						
Investor			0.572+	0.572+				
			(0.339)	(0.338)				
Education Context					3.892***	3.892***		
					(0.557)	(0.557)		
Retail Context							4.896***	4.896***
							(0.654)	(0.654)
_cons	2.570***	2.109***	1.851***	1.269***	3.692***	3.500***	2.478***	2.324***
	(0.245)	(0.126)	(0.200)	(0.094)	(0.298)	(0.168)	(0.201)	(0.113)
N	807	1860	807	1860	807	1860	807	1860
Sample	Original	Full	Original	Full	Original	Full	Original	Full

Notes: This table shows domain experience in the entrepreneurship context. We include subjects across multiple experimental contexts as separate observations. Regressions using the full sample (Columns 2, 4, 6, and 8) include sample (broad vs. original) fixed effects. Robust standard errors are clustered at the subject level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D29: **Domain experience in the hiring context**

	(1) HR Years	(2) HR Years	(3) Data Science Years	(4) Data Science Years	(5) Consulting Years	(6) Consulting Years
Employer	1.192*	1.192*				
	(0.490)	(0.489)				
Data Science Context			3.645***	3.645***		
			(0.434)	(0.434)		
Consulting Context					2.983***	2.983***
					(0.462)	(0.461)
_cons	2.587***	1.847***	1.457***	1.268***	1.522***	1.360***
	(0.229)	(0.112)	(0.143)	(0.085)	(0.149)	(0.085)
N	807	1860	807	1860	807	1860
Sample	Original	Full	Original	Full	Original	Full

Notes: This table shows domain experience in the hiring context. We include subjects across multiple experimental contexts as separate observations. Regressions using the full sample (Columns 2, 4, and 6) include sample (broad vs. original) fixed effects. Robust standard errors are clustered at the subject level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Bénabou, Roland and Jean Tirole, "Incentives and prosocial behavior," *American economic review*, 2006, *96* (5), 1652–1678.

Blei, David M, Andrew Y Ng, and Michael I Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, 2003, *3* (Jan), 993–1022.

Holmström, Bengt, "Managerial incentive problems: A dynamic perspective," *The review of Economic studies*, 1999, *66* (1), 169–182.

Pritchard, Jonathan K, Matthew Stephens, and Peter Donnelly, "Inference of population structure using multilocus genotype data," *Genetics*, 2000, *155* (2), 945–959.