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

Bo Cowgill Pablo Hernández-Lagos Nataliya Langburd Wright*

October 2024

Abstract

Screening human capital based on signals such as job applications or entrepreneurial pitches is crucial for organizations. Signals are informative insofar as they are costly. Generative AI (GAI) complicates screening by lowering the cost of producing impressive signals. We model the informational effects of GAI, showing that applicants' use of 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 the adoption of GAI (ChatGPT) lowers employers' and startup investors' screening accuracy by 4-9%. Consistent with our model, we also find that the use of GAI in some settings improves screening accuracy (in our case, non-English-speaking countries). These results show that GAI can profoundly shape screening accuracy.

Keywords: Screening, Artificial Intelligence, Entrepreneurship, Human Capital.

*All authors contributed equally to this paper. Cowgill: bo.cowgill@gsb.columbia.edu, Hernandez-Lagos: pablo.hernandezlagos@yu.edu, Wright: nlw2136@columbia.edu. We thank Nandil Bhatia, Nan Jia, Ginger Jin, Mike Luca, Neil Thompson, and Emma Wiles for helpful discussion and feedback, as well as participants of the Columbia Management, Analytics, and Data Conference, Conference on Field Experiments in Strategy, and 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 and the Columbia Center for Political Economy. Hernandez-Lagos thanks The 2024 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 to hire the right employees (Wiles et al., 2023; Black et al., 2024) or to invest in promising ideas (Bapna, 2019; Scott et al., 2020). Screening talent requires organizations to evaluate signals generated by the candidates. 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 used to assess human capital (or other states of the world).[1] 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 of screening? In this paper, we develop a novel approach to quantifying *information loss* from GAI. We develop a simple model of how GAI affects learning, and use it to propose a strategy for estimating information loss in 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" to convince employers and investors, respectively, with and without using GAI. We observe each subject's prior experience and expertise, so that we can compare evaluators'

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

assessments of human capital versus the known facts.

On average, the adoption of GAI by candidates causes information losses equivalent to a 4-9% increase in screening errors. When candidates use GAI in our experiments, they produce higher-level signals (higher quality pitches), but their signals become more "compressed" and homogenized, making it harder for evaluators to distinguish underlying skills. We find that when candidates use GAI, evaluators increase their demand for alternative signals by about 6-9%.

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. In some settings, GAI could help low-quality applicants raise their signals, without providing much boost to high-quality ones. In this case, GAI compresses signals between high and low applicants, pools them together, and leaves evaluators unable to separate them (information loss).

In different settings, GAI could help high-quality applicants boost their signals, while providing little benefit on the lower end. Rather than destroying information, we show GAI enhances the signal's informational content in this setting, helping evaluators separate experts and non-experts more effectively (than without GAI). Although this may appear to be a theoretical curiosity, we find evidence in our applied setting (particularly in settings where English is not a first language). For these settings, GAI helps experts significantly more than non-experts. We find that screening accuracy is actually higher when candidates use GAI.

Our findings have implications for talent evaluators, candidates, and GAI companies. Our negative average results suggest that as GAI permeates all forms of media (text, pictures, video, audio, etc.), these signals could become useless for screening.[2] Still, GAI

[2]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; Nosko and Tadelis, 2015; Filippas et al., 2018).

could help screening whenever high-quality candidates improve their signals much more than low-quality ones when using GAI tools. 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.

Our paper contributes to several bodies of literature. The first is about firms' screening and selection processes (Chan and Wang, 2018; Bapna, 2019; Li et al., 2020; Hui et al., 2023; Kokkodis and Ransbotham, 2023; Wright et al., 2023b). Several recent papers examine how evaluators' adoption of AI and machine learning for screening, both for hiring and investing, impact screening outcomes (Horton, 2017; Bhatia and Dushnitsky, 2023). 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 about the use of AI by the candidates themselves (rather than the employer). Several papers suggest that better writing can help matching, both in labor (Hong et al., 2021; Weiss et al., 2022) and product markets (Ghose and Ipeirotis, 2010).[3] Using a field experiment, Wiles et al. (2023) show that writing assistance helps job candidates find employment and finds no evidence that the writing intervention misled employers into expecting higher performance. Gans (2024a) studies how GAI affects communication, emphasizing both salience and signaling mechanisms. We find both information loss and gains in our results, and our model reveals conditions under which these different outcomes emerge.

We also contribute to the literature about the adoption and productivity effects of AI (Agrawal et al., 2021; El Sawy et al., 2021; Allen and Choudhury, 2022; Dell'Acqua et al., 2023a; Gans, 2024b; Kim et al., 2024; Jia et al., 2024). Several recent papers suggest that AI boosts productivity, typically for the firm adopting the AI (Brynjolfsson et al., 2023;

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

Choi and Schwarcz, 2023; Felten et al., 2023; Noy and Zhang, 2023; Eloundou et al., 2024; Svanberg et al., 2024). Our paper raises the possibility of productivity *losses*, arising from outside parties' adopting AI, thereby increasing errors by firms screening these parties.

Within this literature, some papers find that AI is more helpful for firms or workers at the bottom of the performance distribution (more than the top, Brynjolfsson et al., 2023; Dell'Acqua et al., 2023b). Other work suggests the opposite: GAI helps those at the top more (Choudhury et al., 2020; Otis et al., 2023). We contribute a model where this distinction is a parameter. We then show how the accuracy of screening depends on it.

In our empirical section, we find distinct populations based on the geography of the sender (English- versus non-English-speaking countries) where the parameter goes in different directions. We show using experiments that screening accuracy changes with AI in the way our model predicts. Although GAI may seem like a globally-standardized technology, our results suggest that companies in this space may need to pursue context-specific strategies (Bloom et al., 2012; Gupta and Khanna, 2019; Wright et al., 2023a; Carlson, 2023; Wright et al., 2024).

Finally, our study contributes to strategy research about entreprenuership. "Pitches" are a crucial way that venture capitalists screen founders (Clingingsmith and Shane, 2018; Clarke et al., 2019; Falchetti et al., 2022; Gans, 2022). The degradation of signals could create new markets for alternative screening methods in a variety of industries that entrepreneurs could serve (Bingham and Eisenhardt, 2011).

However, signal degradation is not inevitable. Our paper highlights a strategic choice facing GAI startups and incumbents alike: Which customers to prioritize (Gans et al., 2019)? Our paper suggests that in the short run, prioritizing novices could help users "fool" evaluators. As the power of GAI becomes common knowledge, this strategy undermines the signaling value of GAI-created content[4] and could spawn innovation

[4]GAI content may still have value for consumption, but could be less useful for signaling.

in spam-resistant screening. By contrast, GAI developed to help experts (rather than novices) could maintain the informational value of the medium.[5] GAI companies' customer prioritization choices thus have implications for which forms of content or media evaluators use for screening.

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

In this section, we motivate our empirical section using a simple model of how generative AI interacts with screening. Our goal is to understand the long-run impact of GAI. We are particularly interested in a future where GAI is widely used, where gatekeepers are aware of this widespread use, and where they must incorporate its use into inference-making.

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, senders of all types can also use generative AI to enhance their pitch when GAI is available. 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. The receiver faces a signal extraction problem: 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

[5] However, there may be other possible tradeoffs.

accuracy of those beliefs and see how accuracy changes with senders' use of GAI.

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 a cover letter or business plan. Higher signals either require more effort or more expertise to generate. Effort is costly and convex; exerting a effort reduces utility by $a^2/2$. Senders have privately known types in two dimensions:

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 used). We denote this GAI boost as $e_g \in \mathbb{R}$.

These types (e_n, e_g) are privately observed by the sender. 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}. The joint distribution of e_n and e_g is known to all players. As we shall see, the 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 used. 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 use of GAI ($G = 1$ or $G = 0$) is an exogenous, publicly known parameter. We do not model a "choice" to use GAI, but rather compare a pre-GAI era (where $G = 0$, GAI did not exist) and a post-GAI era ($G = 1$). For example, in the $G = 1$ era, GAI may be directly incorporated into word-processing tools, and all senders receive some form of GAI guidance.

Evaluators/receivers in our setup are *aware* that all senders use GAI when GAI is available (e.g., they know that MS Word contains LLM assistance). However, evaluators remain uninformed about any specific candidate's e_g, representing how much GAI helped that specific sender. G is publicly known because all parties know which era they are in.[7]

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 re-

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

ceiver 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 use 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.

2.2 Analysis

We now present the key analytical steps before sharing our main results (which are in Section 2.3). All proofs are in Appendix A. At the core of our results are the receiver's posterior beliefs about the sender's expertise (conditional on the signal s and use 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 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.

Our Lemma A1 result about posterior beliefs allows us to derive two important implications.

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

Appendix A1 establishes exact conditions when the weight $\pi(G)$ increases or decreases. The weight the receiver places on the sender's signal (pitch quality) is important because it influences screening accuracy, as well as 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 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"). An example of this was on display in an infamous federal lawsuit when 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:

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

In Lemma A2, we derive $f_B(b; x, G)$ for our context and simplify the expression for loss. 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 use is $G = 0, 1$. The receiver's average loss is $\mathbb{E}_x[L(x, G)] = (1 - \rho(G)^2)\sigma_n^2$, which equals the variance of receivers' posterior beliefs.*

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 use). 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.

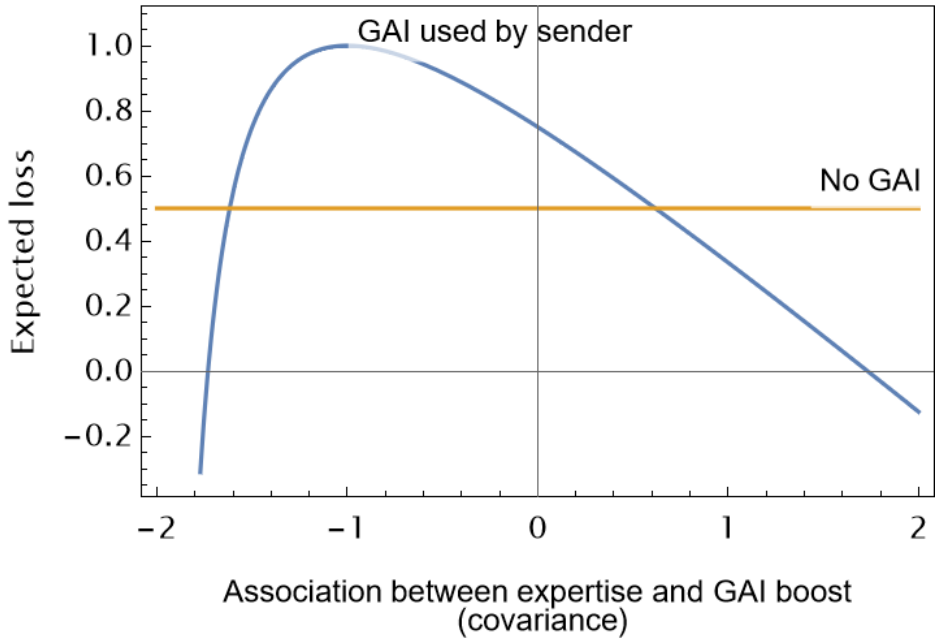
2.3 Main Results

We now have the tools 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 usually extremely higher or extremely lower than a non-expert GAI boost. That is, when the covariance between expertise and GAI boost is extreme.

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

Proposition 1 raises the provocative idea that GAI could actually *improve* the accuracy of screening. It shows that if the covariance σ_{ng} is extremely high (or low), GAI increases the accuracy of screening. Appendix A derives the exact cutoffs for "extremely high" and low (i.e., σ^- and σ^+). The intuition is simple: if the covariance between expertise and GAI boost is extremely high, receivers can easily distinguish experts (experts send very

Figure 1: Information Loss Under GAI



Notes: GAI makes receivers' inferences more accurate when the covariance between expertise and GAI is extreme. Assumptions to draw the plot: All variances are equal to one, except the variance of GAI boost which is set to two.

high signals and non-experts very low ones), and so accuracy improves. However, if the covariance is extremely negative, receivers can *still* distinguish experts (experts are the senders with very low signals), and so accuracy can also improve in this setting.

In other words, for GAI use 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. 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.

This result suggests that in some environments, the opposite of an "epistemic apocolypse" is possibe: better learning, not worse. Are these "extreme" covariances more

than a theoretical possibility? In practice, we suspect that the "extremely low" covariance is less likely.[8] However, extremely high (positive) covariances may be possible.

There is evidence for this possibility in a 2023 experiment in AI art by popular science communicator Cleo Abram (Abram, 2022). 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). Simultaneously, Abram herself (zero art background) completed the same task, developing art both with and without GAI.

The program then had 3,000 subjects blindly evaluate all artwork, including both the professional's and novice's artwork, both with and without GAI. Speaking to the professional, Abram summarized the result:

> *"[The GAI] boosted both of us. By boosting both of us, it gave me new skills, and it* ***gave you superpowers****." (emphasis ours)*

The Abram experiment provides an example of GAI helping an expert more than a novice. 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 (DALL·E) without subject-matter expertise.

The DALL·E experiment was an informal demonstration in a popular science program. Our experiment in this paper scales the Abram experiment under a controlled setting, using the context of hiring and entrepreneurship rather than AI art.

Our final set of results is about the pitch. Does using GAI necessarily make higher signals (i.e., higher-quality pitches)? Does GAI compress the level of signals, homogenizing them, or spread them out? The answer to both questions depends on the covariance σ_{ng}.

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

Proposition 2 (GAI Increases Signal)**.** *GAI increases the average level of the signal 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.*

As with earlier propositions, σ_{ng} increases the level of the signal if and only if it is above a cutoff (derived in Appendix A).

Our final result is about the *dispersion* 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 shows that GAI does not necessarily lead to greater or lower dispersion. As with before, the result is conditional and depends on σ_{ng}. A sufficiently negative covariance between expertise and GAI boost decreases signal dispersion.

3 Empirical Methodology

The main message of our model is: The informational effects of GAI are theoretically ambiguous. The covariance term—capturing the heterogeneity in how GAI affects signals—is a central player in the receiver's ability to learn from the use of GAI. The covariance term may differ across populations and settings.

We now turn to directly measuring the size and direction of informational effects of GAI in practice. The negative epistemic effects of GAI have been much discussed in public discourse. Our experiment aims to 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 measure these effects, we design and execute an experiment outlined below.[9] From the experiment, we can straightforwardly produce reduced-form estimates of how GAI changes signals, beliefs, and accuracy of screening.

Overview of Experiment. We recruit subjects into the role of "senders," tasked with developing a signal (pitch). A second set of "receiver" subjects evaluate the signal and interpret the expertise of the senders. We recruit both sets of subjects from Prolific.[10] In the hiring context, job candidates (senders) write cover letters to persuade employers (receivers) to hire them. In the entrepreneurship context, entrepreneurs (senders) write pitches to investors (receivers) to get funding for their venture idea.

We encouraged senders to write pitches that convey expertise, even if they do not have it. We ask receivers to evaluate the resulting messages and the expertise of the corresponding senders. Specifically, we ask the receivers to assess the overall quality of the pitch and to discern which pitches were written by experts. Critically, we have a measure of whether each sender *actually* has prior expertise in their domain. We can therefore assess the accuracy of the receivers' beliefs about the senders' expertise—both when senders use GAI and not.

Each sender in our design develops four pitches: Two in a domain where they have expertise, and two in a domain where they do not. As such, we have a balanced number of pitches by senders with and without expertise. Within in each domain, each sender first develops a pitch without using ChatGPT. After writing a pitch without ChatGPT assistance, the sender is informed that they can use ChatGPT to rewrite and/or improve

[9] The pre-registration may be found at https://doi.org/10.17605/OSF.IO/RFBU6.

[10] 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.

the pitch. As such, we also have a balanced number of pitches using ChatGPT (and not).

We also recruited receivers with prior experience in the relevant domain to evaluate the senders' pitches. We informed receivers about which pitches used GAI. Each receiver evaluated multiple randomly-assigned pitches. The unit of analysis of our data is thus the combination of sender$\times$receiver$\times$domain$\times$ChatGPT. Because senders produce multiple pitches in different domains, our specifications can use sender fixed effects. Because each receiver evaluates multiple pitches, we can also use receiver fixed effects.

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 entreprenuership) to show 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.

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 openings separately. We task job candidates with developing a pitch for their candidacy for a job opening as would appear, e.g., 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 using ChatGPT.

We ask each receiver to evaluate whether the author of each letter 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 (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 less trust in the pitch as a signal of expertise.

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.

Next, 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. As before, we tell the receivers which of the pitches used ChatGPT. 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. A demand for costly 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 strategy to estimate σ_{ng} from the experimental data, and to conduct statistical tests against the thresholds in Propositions 1-3.

Different populations may have different covariance terms (σ_{ng}), even for the same GAI technology. In our empirical 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, depending on 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. For example, 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 these reasons, the covariance could therefore be positive for senders from outside of an English-speaking background (i.e., ChatGPT will help boost the pitch quality of experts more than 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 experts. Non-experts from an English-speaking background can ask follow-up questions to ChatGPT until their pitch sounds as convincing as 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 sub-

jects 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, occupational, and entrepreneurship experience as provided by this platform. Both senders and receivers get $12/hour for completing the tasks.

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 to not use ChatGPT in the experimental instructions (see Appendix C), some may have done so anyway. To ameliorate 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 directions.[11]

Data. Our final data are on a sender-receiver evaluation level. In total, we have 343 unique senders and 806 unique receivers, about half of each in the entrepreneurship and hiring contexts. Each receiver evaluates eight pitches or cover letters, making our final dataset 6,448 evaluations.[12] Pitches are randomly assigned to receivers within the some context (either startups or hiring). Of our total observations, 3,208 come from the recruitment context, and 3,240 come from the entrepreneurship context. We pool observations across hiring and entrepreneurship in our main paper, and test for heterogeneous effects

[11]For example, some disregarded the prompts all together, i.e., by writing a pitch about the retail industry when the prompt instructed them to do so about the education and training industry.

[12]We lose two sender-receiver observations because of incomplete evaluations, but we recover some of the dependent variables using natural language processing techniques.

in the appendix. The senders come from a mixture of English and non-English-speaking countries.[13]

3.1 Regression Specifications

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

$$Y_{idj} = \beta_1 CHATGPT_{idj} + c_i + r_j + o_{idj} + \epsilon, \tag{3}$$

where Y_{idj} is an outcome for sender i on domain d, evaluated by receiver j. $CHATGPT_{idj}$ reflects whether the sender was using ChatGPT. c_i reflects sender fixed effects, and r_j reflects receiver fixed effects. Because each receiver evaluated multiple senders, we include pitch order fixed effects (o_{idj}).[14] 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, our model shows how 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, and to conduct statistical tests with the cutoffs in Propositions 1-3. In our results section, we mention these estimates of σ_{ng} (and cutoff tests) alongside our reduced-form estimates of our main specification (Equation 3).

3.2 Outcomes

Below we list and motivate all outcome variables. We measure several outcome variables as a 1-5 ordinal rating. In these cases, we use the 1-5 rating as the dependent variable. In

[13] English-speaking countries include: Australia, Canada, South Africa, the United Kingdom, and the United States. Non-English-speaking countries include countries like: France, Mexico, Poland, and South Korea.

[14] For each domain setting, we allocated the pitches across 8 buckets for each ChatGPT condition, so the order fixed effects also account for differences in expert probabilities within each bucket.

some cases, we gave receivers five buckets of choices corresponding to specific numbers or intervals as described below.[15] In these cases, we can either use the 1-5 rating as the dependent variable, or we can use the midpoint of the interval.[16] 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 magnitude 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 (multipicative). The five categories below summarize our outcome variables.

1. **Receivers' beliefs about sender expertise**: We ask receivers to choose a probability bucket representing how likely a sender is to be an expert.
    - 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. **Accuracy**: We measure the accuracy of evaluations in two ways.
    - First, we take the square difference between the actual expertise of senders (binary) and the receiver's probabilistic prediction (mentioned above).
    - Corollary 1 generates an equivalent way to measure accuracy by using the variance of receivers' beliefs. Because receivers' beliefs are probabilities p that the sender is an expert, the variance of this belief is $p \cdot (1 - p)$.

    Both of these outcomes are measures of errors, so positive coefficients indicate less accuracy. Because squared errors and variances are difficult to interpret, we also present

[15] 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.

[16] For example, 10% in our example.

results in logs (for percentage interpretation).

3. **Level of the signal**: We also measure the level of the signal in two ways:

   - First, we ask receivers to rate the pitch directly using a five point scale.
   - In addition, we 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.

     For interpretability, we transform this variable so that higher values represent more sophisticated language.[17] Our NLP variable is 100 minus the raw Flesch-Kinkaid Score.

4. **Dispersion of the signal**: To measure dispersion, we take the mean ratings of the signal by condition (ChatGPT vs. no ChatGPT). We then subtract each rating from the mean within its condition, and square it. The average of these values thus has the interpretation of a variance.[18] We do the same for the NLP rating.

5. **Demand for costlier signals**: Finally, we measure evaluators' demand for a costlier signal than text. 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.

[17] 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. In order to make "higher" scores better, we transformed the raw Flesch-Kinkaid score to our NLP variable equal to 100 minus the raw Flesch-Kinkaid Score.

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

4 Results

Table 1 shows the summary statistics. Many of our main results are in this table, although we later show full regressions with controls. Screening accuracy is lower when senders use ChatGPT by 4-9%. The level of the signal (pitch quality) increases by 3% with ChatGPT when measured as the ratings by receivers. ChatGPT lowers the dispersion of the signal by 10% when measured as the ratings by receivers and by even larger amounts (56-87%) when using the NLP score. On average, receivers are willing to pay 8% more for additional information in the ChatGPT condition than in the control.

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 the regression in Equation 3.

4.1 Screening Accuracy

Table 2 presents our results 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).

Across our outcomes, we find that ChatGPT raises errors. We find this for all of our results in logs, and in levels for the variance of receiver beliefs measure. Our result on the level of squared errors cannot be detected at conventional levels, but is highly significant at conventional levels in percentages/logs. Across our other outcomes, the decreases in accuracy is detectable usual significance levels when using all other outcomes, both with

Table 1: **Summary Statistics by ChatGPT Use**

	With ChatGPT	No ChatGPT	Difference
Sender's True Expertise (0/1)	0.50	0.50	.00016
Receiver's Belief Sender is an Expert (p)	0.45	0.45	.0045
Main Outcomes			
Loss of Screening Accuracy			
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	.0023
Log(Squared Error)	-1.90	-1.99	.089**
Signal (pitch quality)			
Level of Signal (Rating 1-5 by Receiver)	2.91	2.83	.088***
Log(Level of Signal, Rating)	0.96	0.91	.045***
Level of Signal (NLP)	69.96	53.01	17***
Log(NLP Rating)	4.23	3.92	.32***
Signal Variance			
Variance of Rating	1.45	1.59	-.13***
Log(Variance of Rating)	-1.09	-0.53	-.56***
Variance of NLP Rating	133.77	304.70	-171***
Log(Variance of NLP Rating)	3.50	4.37	-.87***
WTP for Costlier Signals			
Receiver WTP	39.40	37.27	2.1***
Receiver WTP (1-5 Rating)	2.47	2.36	.11***
Log(Receiver WTP)	3.39	3.31	.081***

Notes: This table contains summary statistics for ChatGPT and non-ChatGPT messages. $N = 6,446$ total sender$\times$receiver pairs. In $3,277(50.82\%)$ observations, the sender used ChatGPT to develop their pitch. The stars are given by the p-values of the t-test of a difference in means with zero. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

and without fixed effects. The magnitudes are comparable to the ones shown in Table 1, reducing accuracy by 4-9 %.

Why does accuracy go down? In Table D1, we show that ChatGPT does not affect beliefs about the overall level of expertise. Specifically, when senders use ChatGPT, on average, receivers think the senders have about a 45% chance of being an expert. When senders do *not* use ChatGPT, receivers also think that, on average, the sender has a 45% chance of

Table 2: **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 Used in Pitch	.0064***	.0059***	.043***	.04***	.0023	.00078	.089**	.075*
	(.0018)	(.0019)	(.012)	(.012)	(.0067)	(.0074)	(.039)	(.041)
FEs	-	All	-	All	-	All	-	All
R^2	.0023	.26	.0024	.26	.000017	.2	.00084	.22
DV Mean	.17	.17	-1.8	-1.8	.31	.31	-1.9	-1.9
Observations	6,446	6,446	6,446	6,446	6,446	6,446	6,446	6,446

Notes: The table shows results from regressing the measures of screening errors 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.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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 D1 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, ChatGPT makes it more difficult for receivers to separate experts and non-experts when senders are using ChatGPT.

We can interpret our findings in Table D1 through the well-known "bias-variance decomposition."[19] In this classic result, receivers' average squared error equals the receivers' bias (squared) plus their variance. The results of Table D1 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.

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

However, this bias is constant across the two GAI conditions (Table D1). 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 2 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 receiver'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 are easier 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 (e.g., Equation 1). In Appendix Table D2, we show that subjects with prior experience indeed produce pitches rated as higher quality by receivers, both in our ChatGPT sub-sample and not.

Table 3 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 4.5%-6% percent 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, this is a higher increase (+32%) than that for the receivers' ratings of the signal (+4.5-6%). We interpret this as receivers believing there is more to a good pitch than the sophistication of language. When pitching for a job or startup idea, other

Table 3: **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 Used in Pitch	.088***	.12***	.045***	.057***	17***	17***	.32***	.32***
	(.034)	(.032)	(.014)	(.014)	(.41)	(.34)	(.0075)	(.0065)
FEs	-	Full	-	Full	-	Full	-	Full
R^2	.0013	.37	.0019	.37	.25	.7	.26	.7
DV Mean	2.9	2.9	.94	.94	62	62	4.1	4.1
Observations	6,446	6,446	6,446	6,446	6,448	6,448	6,448	6,448

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, receiver, domain, and order fixed effects. Robust standard errors (in parentheses) are clustered at the receiver level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

aspects (e.g., qualifications and ideas) presumably matter a lot as well. 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.

4.3 Dispersion of the Signal

ChatGPT seems to increase the level of signals. But how does it affect the dispersion of the signal? Table 4 shows that ChatGPT homogenizes or "compresses" the level of the signal, lowering the variance between different senders' pitches. Using the receivers' rating of signals (in logs), the variance decreases by 56-58%. Using the NLP measure, the variance decreases by 87-92%. Our signal results fit intuitively with our overall result about screening errors. Screeners are expected to differentiate candidates or entrepreneurs using signals.

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 4: **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	-.13***	-.13***	-.56***	-.58***	-171***	-187***	-.87***	-.92***
	(.041)	(.043)	(.057)	(.063)	(16)	(16)	(.054)	(.053)
FEs	-	All	-	All	-	All	-	All
R^2	.002	.27	.016	.23	.019	.51	.036	.48
DV Mean	1.5	1.5	-.81	-.81	218	218	3.9	3.9
Observations	6,446	6,446	6,446	6,446	6,448	6,448	6,448	6,448

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, 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 Alternative 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 use ChatGPT. Given this, 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 entrepreneurship setting) or market research (in the recruitment setting). Table 5 reports the results. We find that receivers increase their willingness to pay by about 6-9% when senders use ChatGPT regardless of the specification. Consistent with the reduced accuracy of the pitches, this result suggests that receivers find the pitches less informative.

Table 5: **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	2.1***	2.4***	.081***	.089***	.11***	.12***	.055***	.061***
	(.72)	(.62)	(.022)	(.019)	(.036)	(.031)	(.016)	(.014)
FEs	-	All	-	All	-	All	-	All
R^2	.0016	.47	.0024	.46	.0016	.47	.0022	.47
DV Mean	38	38	3.3	3.3	2.4	2.4	.72	.72
Observations	6,446	6,446	6,446	6,446	6,446	6,446	6,446	6,446

Notes: The table shows results from regressing receivers' willingness to pay for costlier signals and belief of the sender being an expert 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.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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

We have thus far shown that ChatGPT lowers screening accuracy, increases signal levels, and lowers dispersion. Our evidence has been average treatment effects from reduced-form specifications. In our theoretical model, these outcomes were a function of the sign and magnitude of a latent parameter σ_{ng}, representing how correlated a sender's expertise is with the signal increase from ChatGPT in a given population.

Studying this parameter empirically can inform the reason underlying our reduced-form results. Propositions 1 through 3 directly connect our main outcomes (accuracy, signal level, and signal dispersion) to this parameter σ_{ng} and its relationship to specific thresholds.

When we apply our strategy for estimating σ_{ng} (Appendix B), our estimate of σ_{ng} is negative—with a point estimate of -0.01 and a bootstrapped standard error of 0.006. Our p-value for testing this estimate against zero is 0.14. The negative value suggests that overall, ChatGPT helped non-experts more than experts.

Our estimate is consistent with the covariance falling between the Proposition 1 intervals, suggesting lower screening accuracy (which is what we find in Table 2). We find that our estimate of the covariance is above the Proposition 2 threshold in which ChatGPT raises the level of the signal. The point estimate is also consistent with ChatGPT increasing the signal dispersion in Proposition 3, although our statistical tests on this parameter are not precise (unlike our reduced form estimates).

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. We can, therefore, estimate the covariance by these sub-populations.

We find that for senders from non-English-speaking countries,[20] σ_{ng} was +0.05 (positive), with a bootstrapped standard error of 0.013, and a p-value of 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 -0.03 with a bootstrapped standard error of 0.007 and a p-value of below 0.001 when tested against zero.

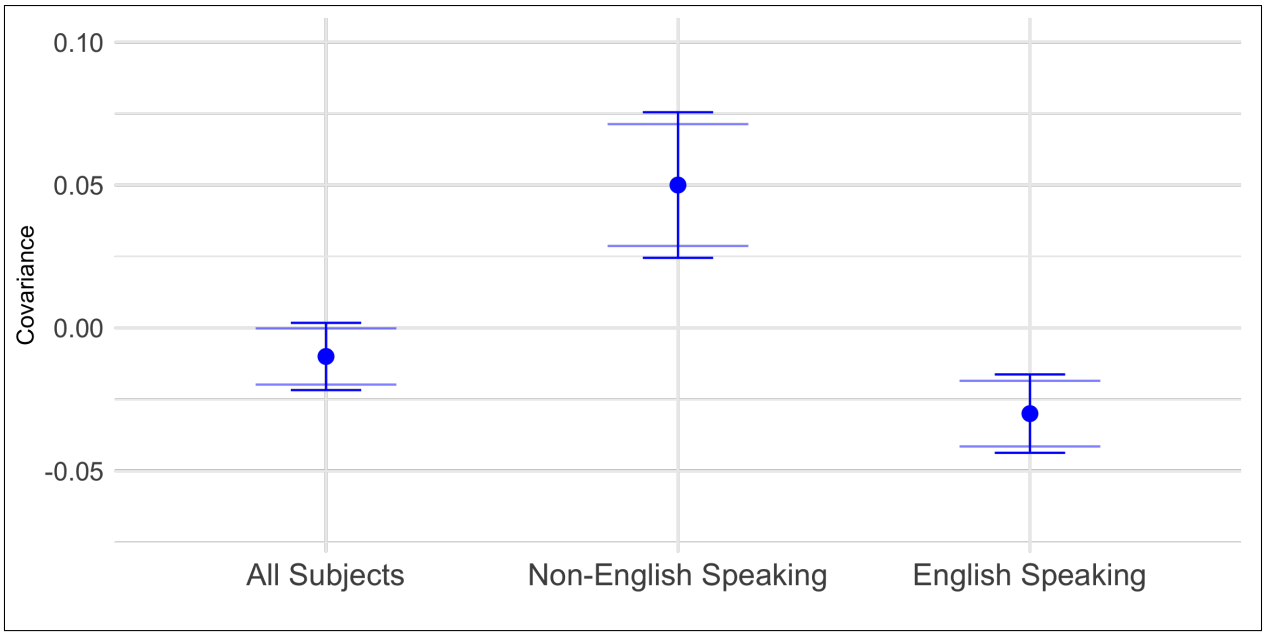
Figure 2 illustrates the differences in our estimate of σ_{ng} across populations. Given that the sign of the covariance differs across these two populations, we now examine heterogeneity in our reduced-form treatment effects of ChatGPT.

4.6 Heterogeneity: ChatGPT and Non-English Sender Contexts

From here, we can now revisit out treatment effects. In Table 6, we study our main outcomes and add an interaction term for non-English-speaking countries. We find that

[20] 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 from 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.

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

However, the impact of ChatGPT on the average level of the signal is lower for senders from non-English-speaking contexts, using the subjective and NLP ratings (columns 3 & 4), which is not consistent with Proposition 2. Finally, the impact of ChatGPT on the dispersion of the signal is higher in magnitude, but not statistically different in non-English-speaking contexts relative to others, qualitatively consistent with Proposition 3 (columns 5-6).

Together, these analyses suggest that the effects of GAI are not monolithic. There are

Table 6: **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
ChatGPT	.16***	.051***	.07***	.34***	-.62***	-.95***
	(.049)	(.014)	(.015)	(.0074)	(.071)	(.062)
ChatGPT× Non-English	-.34***	-.04	-.05*	-.091***	.16	.098
	(.096)	(.026)	(.028)	(.012)	(.13)	(.12)
FEs	All	All	All	All	All	All
R^2	.22	.26	.37	.7	.23	.48
DV Mean	-1.9	-1.8	.94	4.1	-.81	3.9
Observations	6,446	6,446	6,446	6,448	6,446	6,448

Notes: The table shows results from regressing the main dependent variables on ChatGPT and ChatGPT interacted with whether the sender is from a non-English-speaking country. 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.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

cases when GAI can actually increase accuracy in screening. Consistent with the model, when experts benefit more than non-experts, as we see among senders in non-English-speaking contexts, ChatGPT actually helps evaluators screen candidates.

5 Discussion and Conclusion

Screening is abundant in modern life and 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 GAI could profoundly change how these agents learn and screen. Many signals are useful for evaluators because of their costliness. If GAI makes these signals cheaper, their informational content changes.

We are not the first to raise the possibility of information loss. Horton (2024) writes, "the problem of deciding what's real or not will become acute," and describes it as "a very real problem we all now have." Concern about fake news, cheating on homework,

and identity fraud—all examples of information loss—abound in discourse around GAI.

Our paper aims to contribute a deeper conceptual understanding of GAI information loss and high-quality 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 (minus 4-9%) as well as the magnitude of the parameters generating that change (a negative $\sigma_{ng} = -0.01$).

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 increase the accuracy of screening. While these instances may be relatively rare in the current environment, GAI companies could shape this outcome in future. These companies have some control over whether their products are more helpful for novices or experts (σ_{ng}), and thus over the signaling value of the content that their software produces.

Our results suggest a few lines of future research. 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 could improve 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.

The nature of alternative, spam-proof screening mechanisms is another exciting topic of future research. Horton (2024) suggests the learning problems discussed in this paper could be addressed with cryptography. Other commentators have suggested a greater role for live in-person assessments without the use of AI assistance. In some cases, the overall need for screening may actually decrease. If an employee can use GAI on the job, it might not matter how productive they are without it. These issues raise questions about

managerial best practices for organizations and entrepreneurs working with AI.

If text and other signals becomes unreliable, better signals could be more expensive. Will screeners turn to nuanced investigations of individual quality; for example, through a full background investigation or market analysis? Or will receivers instead rely on observable characteristics of individuals that are easy to verify, such as gender, race, or geography? Understanding how generative AI impacts screening has important implications for accuracy and bias in hiring, entrepreneurial finance, and many other economic transactions. We hope that future work investigates these possibilities.

References

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

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.

Bingham, Christopher B. and Kathleen M. Eisenhardt, "Rational heuristics: the 'simple rules' that strategists learn from process experience," *Strategic Management Journal*, 2011.

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

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

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

Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond, "Generative AI at work," *Working Paper*, 2023.

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

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.

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.

Clarke, Jean S., Joep P. Cornelissen, and Mark P. Healey, "Actions speak louder than words: How figurative language and gesturing in entrepreneurial pitches influences investment judgments," *Academy of Management Journal*, 2019.

Clingingsmith, David and Scott Shane, "Training aspiring entrepreneurs to pitch experienced investors: Evidence from a field experiment in the United States," *Management Science*, 2018.

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

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.

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

Falchetti, Denise, Gino Cattani, and Simone Ferriani, "Start with "why," but only if you have to: The strategic framing of novel ideas across different audiences," *Strategic Management Journal*, 2022.

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., "A theory of visionary disruption," *NBER Working Paper No. w30091*, 2022.

_ , "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.

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.

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.

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

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

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

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.

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.

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.

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

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

_ , _ , and _ , "Contributing to Growth? The Role of Open Source Software for Global Startups," *Harvard Business School Working Paper 24-040 Available at SSRN 4699182*, 2024.

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

Appendix: For Online Publication Only

A Theory Appendix

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

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

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. Lemma A1.

This result follows Bénabou and Tirole (2006). 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 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 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. 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 yield 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 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 $\sigma_{ng} \in [-\sigma_n\sigma_g, \sigma_n\sigma_g]$. For example, if σ_ϵ is close to zero, whenever σ_{ng} is sufficiently below $-\sigma_g^2$, GAI increases the weight. Otherwise, GAI decreases it. □

Proof. Corollary 2. Optimal effort in equilibrium is $a^* = r\pi(G)$ (Equation 6), where $r > 0$ constant. The result follows from Corollary 1. □

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 deviation from average expertise and the variance of expertise.

Proof. Lemma A2.

Loss measures the receivers' beliefs' expected deviations when the sender's true expertise is $e_n = x$:

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

where $\mathcal{B} \equiv E[e_n|s-a^*,G]$ is a random variable that represents the receivers' beliefs. 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 $E[\mathcal{B}|e_n = x, G]$:

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

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, and

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

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 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 as the correlation between beliefs and expertise becomes extreme: 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:

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

Loss is therefore

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

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

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

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

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

The expressions for the squared correlations are:

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

Manipulating these expressions, 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

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

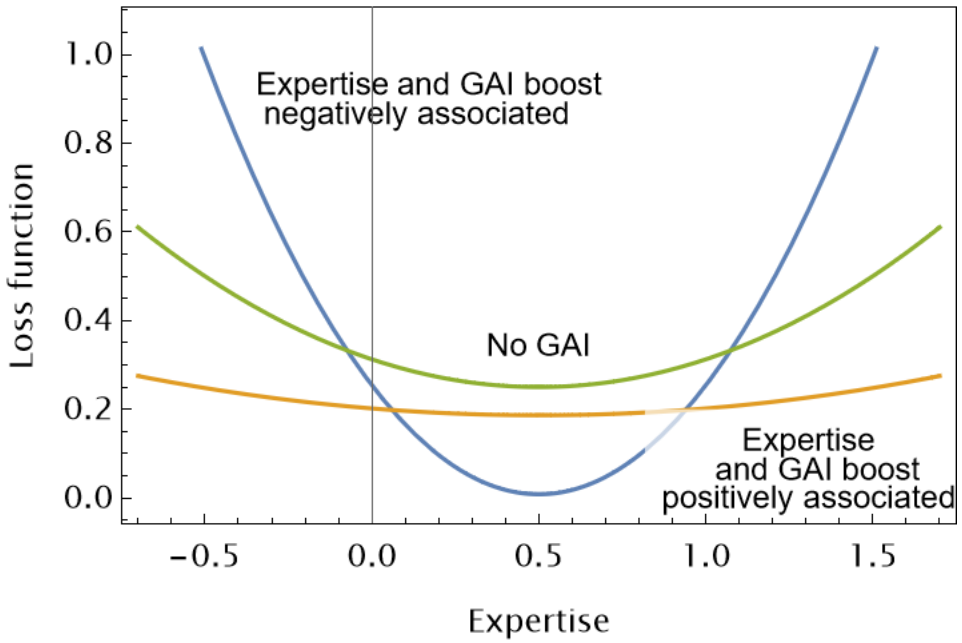


Figure A1: Loss for different covariance values between expertise and GAI boost. All other noise parameters are normalized to one, and the mean expertise is set to 0.5. For example, when the covariance between expertise and GAI boost is negative (set to -0.9 in the figure), the loss of screening extreme 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.

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. Proposition 2. Define $s^*(G) = s - a^*$ for $G = 0, 1$. We need to establish that $s^*(1) - s^*(0) \geq 0$, which is equivalent to $\pi(1) - \pi(0) \geq -\bar{e}_g/r$. 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. □

Proof. Proposition 3. It 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 to 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 use $G = 0, 1$. The receiver puts more weight on the signal when GAI is used 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. 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 $\pi(1) > \pi(0)$. □

The last result highlights the difficulties in assessing expertise when GAI is used. First, using GAI can be *more* informative than not using it 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 used than when GAI is not used. In contrast, when the signal perfectly reveals the sender's expertise without GAI, $\sigma_\epsilon^2 = 0$, a negative, large-magnitude covariance makes using GAI more informative than not using it. Second, in cases where noise is small or zero, when experts benefit from GAI more than non-experts (positive covariance), using GAI makes using the signal less informative than not using it. The intuition for this result is that a positive covariance adds more noise to the signal than information, impairing the receiver's inference. A corollary of Proposition A3 is that a covariance close enough to zero makes the signal less informative when GAI is used 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.

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, and for conducting statistical tests of our estimate against the cutoffs in our propositions. 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.*

The quantity Δ can be measured empirically for each sender by measuring pitches with and without 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. 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 use). 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}. □

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 used ChatGPT in one observation and did 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 -0.01, with a bootstrapped a standard error of 0.006. Our p-value for testing this estimate against zero is 0.14.

For non-English-speaking sender contexts, σ_{ng} was +0.05, with a bootstrapped standard error of .013, and a p-value of below 0.001 when tested against zero.

For senders who *do* come from English-speaking contexts, σ_{ng} was -0.03, with a bootstrapped standard error of .007, and a p-value of below 0.001 when tested against zero.

Lemma B4 (Accuracy)**.** *Let σ_{ng} be the covariance between expertise and GAI boost, and $\rho(G)$ the correlation between expertise and the signal when GAI use is $G = 0, 1$. GAI increases receivers' average accuracy if and only if one of the following two equivalent conditions is met:*

1. *σ_{ng} falls outside the interval (σ^-, σ^+) for constants $\sigma^- < 0$ and $\sigma^+ > 0$.*
2. *The correlation squared when GAI is used, $\rho(1)^2$, is greater than the correlation squared when GAI is not used, $\rho(0)^2$.*

GAI strictly decreases accuracy otherwise.

Proof. Proposition B4. Let us start with the second item in the proposition. From the expressions for the expected loss

$$\begin{aligned}\mathbb{E}_x[L(x,1)] &= (1-\rho(1)^2)\sigma_n^2 \\ &= (1 - \left(\frac{\sigma_n^2 + \sigma_{ng}}{\sigma_n\sqrt{\sigma_n^2 + 2\sigma_{ng} + \sigma_g + \sigma_\epsilon^2}}\right)^2)\sigma_n^2,\end{aligned}$$

where $\rho(G)$ when $G = 1$ is the correlation coefficient between expertise and the signal in equilibrium as given in Lemma A2. Likewise,

$$\begin{aligned}\mathbb{E}_x[L(x,0)] &= (1-\rho(0)^2)\sigma_n^2 \\ &= (1-\left(\frac{\sigma_n^2}{\sigma_n\sqrt{\sigma_n^2+\sigma_\epsilon^2}}\right)^2)\sigma_n^2.\end{aligned}$$

From the first lines in each of these expressions

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

For the first item in the proposition, we compare the squared correlations as before, isolating the covariance between expertise and GAI boost from the other parameters:

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

for GAI to increase accuracy (reduce expected loss). Manipulating these expressions, some 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:

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

□

Although our covariance is negative and close to zero, we need to make sure it is not lower than σ^- in the Proposition 1. Using Lemma B4, we can directly compute the correlations between the signal level (receiver rating) and expertise when GAI is used, and separately when it is not used. We call this $\rho_{G=1}$ and $\rho_{G=0}$. Because the correlations are positive, accuracy decreases with GAI if we can reject the null that $\rho_{G=1} - \rho_{G=0} \geq 0$. We bootstrapped the distribution of this difference. We cannot reject that this difference is equal to zero at conventional levels. However, the mean of the difference between the correlations is -0.003, with a bootstrapped standard error of +0.02. The negative mean suggests that accuracy would decrease.

We now turn to the thresholds in Proposition 2 that determine if ChatGPT increases the level of the signal. We use the following result:

Lemma B5. *Let $\bar{e}_g \geq r > 0$, so the mean GAI boost is not too small. Let $s^*(G)$ denote the signal ("pitch" quality) when GAI use is $G = 0, 1$. GAI increases the average level of the signal if and only if one of the following equivalent conditions hold:*

1. *$\sigma_{ng} \geq \sigma^S$, where σ^S is a negative constant, or*
2. *$\frac{Cov(e_n, s^*(1))}{Var(s^*(1))} - \frac{Cov(e_n, s^*(0))}{Var(s^*(0))} \geq -\bar{e}_g / r$.*

and strictly decreases it otherwise.

Proof. Lemma B5. Define $s^*(G) = s - a^*$ for $G = 0, 1$. We need to establish that $s^*(1) - s^*(0) \geq 0$, which is equivalent to $\pi(1) - \pi(0) \geq -\bar{e}_g / r$. The first item in the proposition comes from some 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. The second item follows from the definition of $\pi(G)$:

$$\pi(1) - \pi(0) \geq -\bar{e}_g / r \iff \frac{Cov(e_n, s^*(1))}{Var(s^*(1))} - \frac{Cov(e_n, s^*(0))}{Var(s^*(0))} \geq -\bar{e}_g / r.$$

□

Lemma B5 states that comparing the covariance between expertise and GAI boost is equivalent to comparing coefficients from linear regressions. The first term on the left-hand side of condition #2 in the lemma is the coefficient on the signal when we regress expertise on the signal in the case that senders used GAI. The second term is the coefficient on the same regression but estimated over the no-GAI condition.

We reject the hypothesis that the LHS is less than the RHS. We computed this using a regression predicting expertise using the ratings and the interaction of ChatGPT and ratings. The difference in the left-hand side of #2 in the lemma is the coefficient in the

interaction from this regression. We then compare this interaction coefficient to the mean of Δ as a proxy for $-\bar{e}_g/r$. We reject that the coefficient on the interaction is less than the mean of Δ, and the p-value is less than 0.001. This suggests that the signal level will increase with ChatGPT use.

Finally, to test whether GAI increases the variance of the signal, it suffices to run a regression of expertise and GAI boost (measured as Δ in Definition 1). A coefficient smaller than $-1/2$ implies that GAI reduced the variance of the signal. When we run this regression on our data, we find a coefficient of -0.01 with a standard error of 0.03. We reject that this coefficient less than -1/2, although we do find in our reduced form estimates that the signal variance lowers with ChatGPT.

C Subject Instructions

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 Receiver

Below we include instructions that we gave to investors in the education and training industry in the entrepreneurship context. The instructions are similar as for receivers

in the retail industry in the entrepreneurship context, as well as for those with data science/management consulting backgrounds in the hiring context.

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.

C3 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,

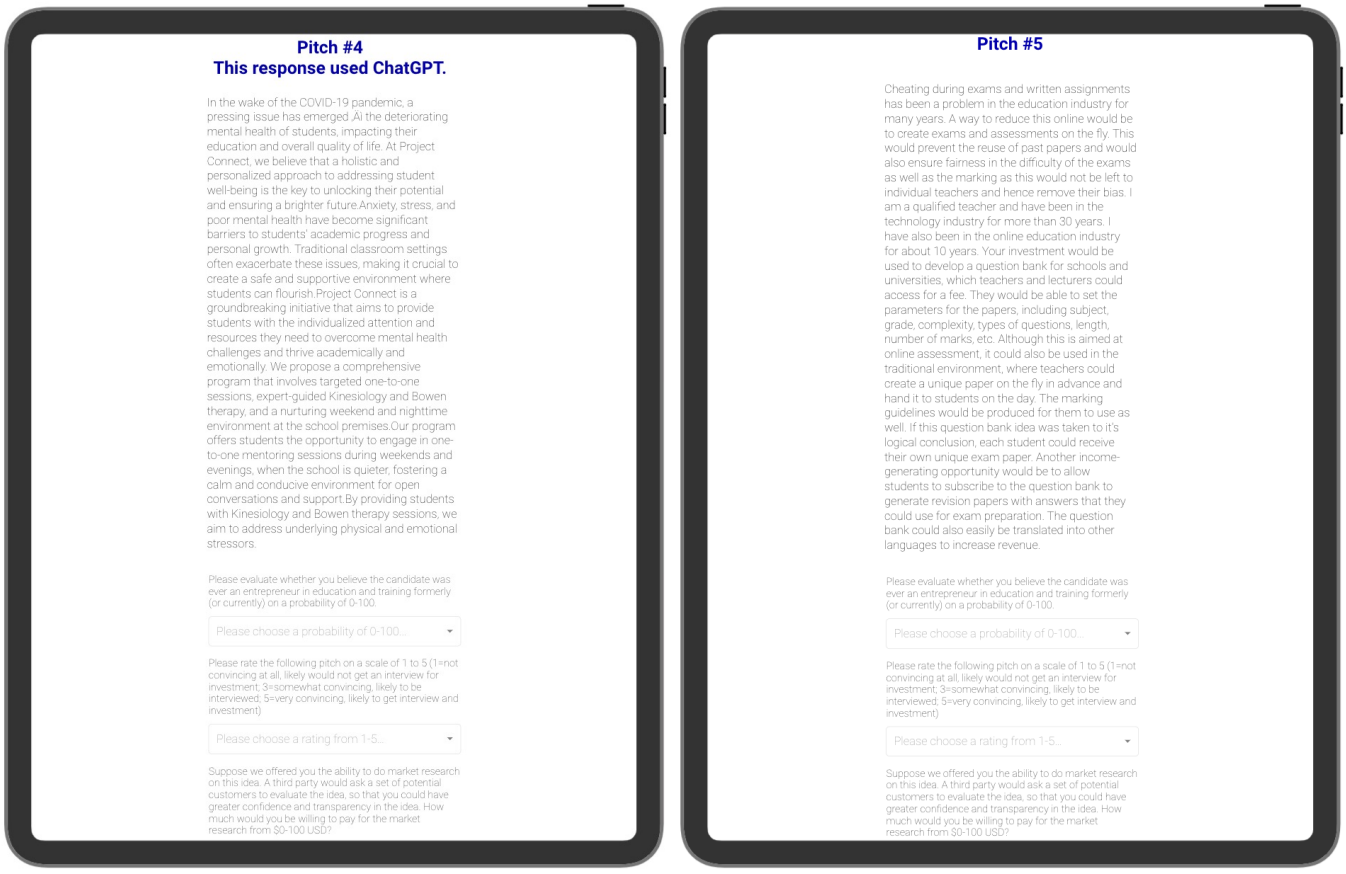


Figure C2

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).

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 C2 illustrates this evaluation task. The pitch on the left uses ChatGPT, while the one on the right does not. Each receiver evaluates eight of such screens.

D Additional empirical results

Table D1: **ChatGPT and Beliefs of Expertise**

	(1) Pr Sender is Expert	(2) Pr Sender is Expert
ChatGPT Used in Pitch	.0045	.0077
	(.0076)	(.0073)
FEs	-	All
R^2	.000068	.41
DV Mean	.45	.45
Observations	6,446	6,446

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

Table D2: **Expert Signal Quality is Higher**

	(1) Signal Level (Rating)	(2) Signal Level (Rating)	(3) Signal Level (Rating)
Expertise	.16***	.13***	.17***
	(.031)	(.045)	(.046)
FEs	All	All	All
R^2	.37	.52	.56
DV Mean	2.9	2.9	2.8
Observations	6,446	3,246	3,134

Notes: The table shows results from regressing the level of the signal on whether senders are experts. Robust standard errors (in parentheses) are clustered at the receiver level.
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D1 Difference between Hiring and Entrepreneurship Contexts

Table D3: **Summary Statistics by Setting**

	Startup	Hiring	Difference
Sender's True Expertise (0/1)	0.49	0.50	-.0083
Receiver's Belief Sender is an Expert (p)	0.43	0.47	-.034***
Variance of Receiver Belief, $p \cdot (1-p)$	0.17	0.17	.00033
Log(Variance of Receiver Belief)	-1.84	-1.84	.00089
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	2.97	-.19***
Log(Level of Signal, Rating)	0.90	0.98	-.083***
Level of Signal (NLP)	59.76	63.51	-3.7***
Log(NLP Rating)	4.03	4.12	-.089***
Variance of Rating	1.55	1.48	.068*
Log(Variance of Rating)	-0.76	-0.86	.095*
Variance of NLP Rating	247.68	187.72	60***
Log(Variance of NLP Rating)	4.05	3.81	.23***
Receiver WTP	38.59	38.11	.48
Receiver WTP (1-5 Rating)	2.43	2.41	.024
Log(Receiver WTP)	3.36	3.34	.012

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 D4: **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	-.0083	.0028	.091**	12***	-.12*	-166***
	(.01)	(.0028)	(.045)	(.36)	(.063)	(15)
ChatGPT$\times$ Startup	.018	.0062*	.054	9.3***	-.026	-41
	(.015)	(.0037)	(.063)	(.59)	(.086)	(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	6,446	6,446	6,446	6,448	6,446	6,448

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

References

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

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