Labor Market Signals: The Role of Large Language Models*

Kian Abbas Nejad Giuseppe Musillo

Till Wicker Niccolò Zaccaria

Abstract

Large Language Models (LLMs) are transforming the labor market, including hiring decisions. This paper examines their impact on the signals job-seekers send to potential employers through two field experiments focusing on cover letters. We find that LLMs enhance signal quality, especially benefiting lower-quality applicants, yet these gains do not boost interview invitations because they are concentrated in standardized, less influential sections of the cover letter. When recruiters learn of LLM usage, they place greater value on high-quality, human-crafted letters. Hence, LLMs reduce the informativeness of signals, potentially increasing inefficiencies in labor market matching as estimated by a calibrated structural model.

Key words: Large Language Models; Cover Letters; Labor Market; Matching; Signaling.

JEL codes: C93; J24; O33; M51

*[**Abbas Nejad**: Tilburg University (e-mail: k.abbasnejad@tilburguniversity.edu); **Musillo**: Tilburg University (e-mail: g.musillo@tilburguniversity.edu); **Wicker**: (e-mail: t.n.wicker@tilburguniversity.edu) **Zaccaria**: (e-mail: n.zaccaria@tilburguniversity.edu). All authors contributed equally. We thank Francesco Capozza, Patricio Dalton, Mery Ferrando, Matteo Giugovaz, Djurre Holtrop, Boris van Leeuwen, Melika Liporace, Janneke Oostrom, David Schindler, Sjak Smulders, Has van Vlokhoven, Anna Salomons, Sigrid Suetens, and Burak Uras for their valuable comments. We also thank participants at mEETing in Tilburg University and the University of Tokyo for helpful discussions. Financial support from the Economics Department of Tilburg University is gratefully acknowledged. We thank OpenAI for supporting the textual analysis through its Researcher Access Program. The experiments received IRB approval (IRB FUL 2024-002 and TiSEM_RP1776, respectively) and were pre-registered (AEARCTR-0013355 and AEARCTR-0014784, respectively).

1 Introduction

The labor market is undergoing rapid transformations, driven by technological innovation and evolving workplace practices. Recent research highlights the impacts of these developments on modern labor dynamics (Deming and Kahn, 2018; Englmaier et al., 2024; Deming et al., 2025). Among the emerging technologies, the rise of Large Language Models (LLMs) promises to have substantial implications for the future of work, with at least 80% of the U.S. jobs expected to be affected, and 23% of U.S. employees already using LLMs at work (The White House, 2022; Eloundou et al., 2023; Bick et al., 2025). However, relatively little attention has been given to how LLMs can influence labor market matching: the process through which job-seekers are matched to firms.

Labor market matching heavily depends on signals, such as CVs and cover letters, since firms cannot directly assess a job-seeker's productivity or fit. Cover letters, in particular, play a crucial role in many hiring processes, allowing job-seekers to showcase soft skills like motivation, communication abilities, and overall fit (ANP, 2023, 2024). 60-89% of recruiters across the US and the Netherlands require job-seekers to submit a cover letter, but the impact of LLMs on these signals is not clear ex-ante.[1] On one hand, LLMs may enhance job-seekers' signals by improving writing quality (Noy and Zhang, 2023). On the other hand, they might render texts more formulaic and less personalized (Shanahan, 2024), reducing signal effectiveness. Despite uncertainty about whether LLMs help or harm employment prospects, more than half of job-seekers use them in applications (Criddle and Strauss, 2024). Do labor market signals written with LLM assistance help job-seekers get hired, and does this distort hiring decisions?

This paper answers this question through two pre-registered field experiments examining the impact of LLMs on labor market signals and analyzing both the perspectives of job-seekers and employers. The first experiment is conducted with job-seekers at two universities and recruiters from four multi-national companies with close to half a million employees. During the field experiment, some job-seekers were allowed to use ChatGPT

[1] See Rendement (2024), Resume Genius (2024), and Zety (2024). The four most commonly cited pieces of information obtained from cover letters are: motivation for the job, language usage, match with the organization, and match with the job. Recruitment expert Allison Hemming has even argued that modern screening technologies (such as those evaluated by Avery et al. (2024); Chaturvedi and Chaturvedi (2025); Awuah et al. (2025)) give recruiters more time to read cover letters, hence increasing their importance (Schwartzberg, 2025).

when writing their cover letter, while others were not. Access to LLMs improves the quality of cover letters by more than 0.2 standard deviations, as evaluated by recruiters from the four firms who were blind to the two treatments. In line with Noy and Zhang (2023), Dell'Acqua et al. (2023), Caplin et al. (2024), and Brynjolfsson et al. (2025), the positive treatment effects are stronger for lower-quality job-seekers, thus reducing the dispersion in quality of the cover letters, hence reducing their value as a signal of the job-seeker's true ability.

However, having access to LLMs does not increase the job-seeker's likelihood of advancing to the next recruitment stage, an interview. This is because recruiters primarily care about two dimensions of the cover letter, neither of which improves as a result of access to LLMs. The first dimension is the job-seeker's motivation. We find that ChatGPT improves less personalized and more generic sections, but falls short in enhancing more distinctive and individualized components such as the job-seeker's motivation. Textual analysis of ChatGPT conversation histories reveals that job-seekers ask ChatGPT for help with their motivation multiple times; however, ChatGPT does not significantly improve this section.

The second dimension recruiters care about is the cover letter's clarity, akin to recruiters in Wiles et al. (2025). Access to LLMs does not affect the perceived clarity of the cover letter (as evaluated by the recruiter) due to two competing forces. While LLMs enhance the cover letter's clarity by reducing the frequency of grammatical mistakes by 39%, LLMs reduce clarity by *worsening* the text's readability through using longer words. These two objective effects go in opposite directions, with LLMs not affecting the recruiter's perception of the cover letters' clarity. The lack of improvements in the motivation and clarity of the cover letter explains why, although LLMs improve the average quality of the cover letter, they do not increase the likelihood of the job-seeker being invited to an interview, the next stage in the recruitment process.

Recruiters' perceptions of LLM usage can also affect the evaluation of labor market signals, with people having a preference for human over LLM-generated content (Zhang and Gosline, 2023). To understand this, we conducted a second experiment where we employed 401 recruiters to evaluate a subsample of cover letters from the first experiment. Some were explicitly informed about whether a cover letter had been written with LLM assistance, while others were only told that some letters had been written with the assistance of LLMs without knowing which ones. When recruiters were explicitly informed about LLM usage, they do not rate cover letters differently. However, there is substantial heterogeneity: while the evaluation of low and medium-quality cover letters do not differ systematically, recruiters evaluate high-quality cover letters written without LLM assistance more positively and are more likely to invite the applicant to the next stage of the recruitment process.

Related Literature Several studies have evaluated the role of technologies on employee-employer matches in the labor market, including the internet (Autor, 2001) and algorithmic writing assistants like Grammarly (Wiles et al., 2025). However, to the best of our knowledge, this is the first study to provide causal evidence on how the use of LLMs by job-seekers influences signaling and matching outcomes. Unlike aforementioned technologies, LLMs can generate tailored job application materials, fundamentally changing how job-seekers present themselves to employers. As these tools become increasingly widespread, their use among job-seekers is likely to have a greater impact on the matching efficiency in the labor market.

The study most closely related to ours is Wiles et al. (2025), who randomize access to an algorithmic writing assistant (AWA) to job-seekers on an online gig platform. Unlike LLMs, the algorithmic writing assistant studied in Wiles et al. (2025) provides suggestions on text already written by the job-seeker, while LLMs can do this and generate text themselves. Importantly, our study focuses on entry-level positions in full-time employment—jobs where soft skills, often conveyed in cover letters, play a central role in selection (Deming, 2017). In contrast, the gig-economy context analyzed by Wiles et al. (2025) places less emphasis on soft skills, and involves shorter, resume-style documents with limited narrative content.[2,3]

[2] Less than 1% of U.S. workers were employed in online gig roles in 2017 versus 71% in full-time jobs by 2022 (United States. Bureau of the Census and United States. Bureau of Labor Statistics, 2021; U.S. Bureau of Labor Statistics, 2023). When hiring for entry-level positions, signals are scarce because of limited past work experience, yet stakes are high due to the path dependency of future jobs (Arellano-Bover, 2024).

[3] Our setting differs from Wiles et al. (2025) along several other important dimensions. Firstly, the documents evaluated differ: in Wiles et al. (2025), treated job-seekers had access to AWA when writing a 70-word resume focused on their professional experience. Instead, in our study, job-seekers wrote cover letters that averaged 430 words and focused primarily on their motivation and soft skills. Secondly, the timing differs: Wiles et al. (2025) was conducted in 2021, before the emergence of LLMs, and hence recruiters took signals at face value. In contrast, our study was conducted in spring 2024, when LLM adoption among job-seekers was already widespread and recruiters were aware of this (Criddle and Strauss, 2024). Lastly, the sample differed substantially: in Wiles et al. (2025), only 20% of the sample comes from anglophone countries, with the majority of job-seekers' first language not being English. Instead, our study consisted of highly educated individuals who had completed at least a Bachelor's degree with English as the language of instruction. This is reflected in the spelling and grammar rate in the control group, which was 8.0% in Wiles et al. (2025), and 2.8% in our study.

This difference in job type and document structure implies different signaling environments and raises distinct questions about how job-seeker's signals are interpreted by employers.

To interpret our results, we adopt the "clarity" versus "signaling" framework introduced by Wiles et al. (2025). In this framework, clarity refers to the extent by which writing reduces information frictions by making a job-seeker's true attributes easier to assess—for example, by correcting grammatical errors or improving structure. Signaling, by contrast, captures the idea that the information conveyed in a cover letter can serve as a proxy for applicant ability: well-written applications credibly indicate competence, effort, or communication skills. In Wiles et al. (2025), AWAs improve clarity by reducing the number of grammatical and spelling mistakes and improving the signal's readability, but leave the underlying signal untouched. In contrast, we find a more nuanced story: LLMs on net do not affect the clarity of the cover letter, but affect the conveyed signal. Further, LLMs improve the cover letter's components that are not important in the recruiter's decision to interview a candidate or not.

How these changes in labor market signals ultimately affect hiring depends in part on how recruiters perceive and respond to them. Findings from our second experiment shed light on this, revealing that recruiters' perceptions—thus far overlooked in the literature—play a central role in the interaction between LLM-usage and hiring decisions. While existing studies have looked at the role of AI- and LLM-generated recommendations for evaluators (Hoffman et al., 2017; Dell'Acqua, 2022; Avery et al., 2024; Chaturvedi and Chaturvedi, 2025; Awuah et al., 2025), our study looks at how evaluators perceive and judge written content created by job-seekers with and without access to LLMs. Given the widespread use of LLMs by job-seekers, this is an important distinction (Criddle and Strauss, 2024). Our second experiment also contributes to the literature on the evaluation of human vs. LLM-generated content (Böhm et al., 2023; Noy and Zhang, 2023; Zhang and Gosline, 2023; Kadoma et al., 2024; Bohren et al., 2024) by illustrating the implications on the labor market, and by highlighting the role that perceptions of the evaluator play. Furthermore, we show that recruiters are no better than chance at correctly identifying whether a text was written with or without LLM access.

By substituting for job-seeker effort and distorting the informational content of applications, LLMs introduce a novel form of mismatch in the labor market. Unlike non-generative AI tools such as AWAs, LLMs not only reduce the discrepancy in the quality of written applications, but also blur the link between the observed signal and the job-seeker's underlying ability. This decoupling raises important questions about how hiring decisions are made in equilibrium. To study the broader implications of these distortions, we develop an assignment model with imperfect information, calibrated using findings from our experiments. Building on the existing literatures on labor market signaling (Spence, 1973; Kurlat and Scheuer, 2020; Bassi and Nansamba, 2021; Carranza et al., 2022) and matching (Teulings, 1995; Eeckhout and Kircher, 2018), we model how LLMs alter signal quality, incorporating experimentally observed asymmetries whereby LLMs disproportionately inflate signals from lower-ability applicants. Our calibration exercise quantifies substantial welfare losses resulting from these distortions, estimating efficiency losses of up to 0.8% relative to a counterfactual scenario with no LLM adoption among job-seekers. These losses emerge because firms, anticipating signal inflation, discount all applications and make suboptimal hiring decisions.

Lastly, our paper contributes to the growing body of evidence on the effects of generative AI and LLMs on productivity across a range of domains, including professional writing tasks (Noy and Zhang, 2023), law school exams (Choi and Schwarcz, 2024), coding tasks (Peng et al., 2023b), consulting (Dell'Acqua et al., 2023), customer support (Brynjolfsson et al., 2025), and business practices (Otis et al., 2024).[4] Unlike these settings, we focus on a personalized and persuasive writing task. This is an important distinction from existing studies, as LLMs may struggle with personalized persuasive writing due to their formulaic nature. Our results confirm this, as LLMs greatly improve non-personalized sections of the cover letter (e.g., introduction, conclusion), but do not improve the personalized sections (e.g., motivation). Textual analysis of conversations between the job-seeker and ChatGPT reveals that job-seekers repeatedly ask ChatGPT to improve the personalized sections, to no avail. In turn, ChatGPT prompts substitute google searches, resulting in fewer overall and grammar-related searches. We also provide novel evidence that while LLMs correct spelling and grammar mistakes, they simultaneously reduce the text's readability by increasing average word length.

The remainder of this paper is structured as follows. Sections 2 and 3 discuss the

[4] Several other papers have studied LLMs and their implications. A non-exhaustive list includes Peng et al. (2023a); Acemoglu (2024); Autor (2024); Merali (2024); Carvajal et al. (2024); Capraro et al. (2024); Filippucci et al. (2024); Cui et al. (2025).

experimental design and results of the job-seeker experiment, while Section 4 describes the recruiter experiment. Section 5 presents the assignment model and quantification exercise, before Section 6 concludes.

2 Job-seeker Experiment: Design

To understand what effect LLMs have on cover letter quality, we recruited 137 students from two of the largest universities in the Netherlands, Tilburg and Utrecht University, to participate in a field experiment with four multi-national corporations: Philips, PwC, Rabobank, and VodafoneZiggo. The firms have a global workforce of over 480,000, and offer highly coveted positions. Each firm provided us with an entry-level job description, and two recruiters from their HR division.

During a university-wide career week, students signed up to our *Cover Letter Challenge*, by uploading a CV, filling in basic demographic information (which were used to stratify randomization), and indicating their preferred firm.[5] The event consisted of a 10-minute introduction, followed by 1 hour to write a cover letter to their preferred firm.[6] Job-seekers were informed that their cover letter and CV would be pseudonymized, and subsequently shared with two independent recruiters of the firm they applied to. Recruiters evaluated the job-seeker's pseudonymized CV and Cover Letter against evaluation criteria co-developed by the researchers in collaboration with the firms. Recruiters only evaluated application packages submitted to their firm, and hence received application packages by job-seekers in both the control and treatment group—however they were blind to the two treatments.[7] Recruiters evaluated five dimensions of the CV and cover letter individually, before assigning a final grade to the CV, cover letter, and complete application package. They subsequently indicated the likelihood of inviting the applicant to a job interview on a 5-point Likert scale.

[5] We focused on students who were in their last semester of either their Bachelor's or Master's degree, and who were actively looking for a job.

[6] Although job-seekers were told they only had a full hour to complete their cover letters, in reality the time constraint was not binding. Nevertheless, 83% finished their cover letter within 60 minutes.

[7] See Appendix B.1.3 for the evaluation criteria. In line with the evaluation criteria, recruiters used cover letters to understand an applicant's motivation, soft skills, and overall fit. These components play a large role for companies, as documented in the desired characteristics listed in the firm's job description. Among others: "Ambitious, proactive, analytical, good with figures, results-focused and flexible"; "Excellent communication and influencing skills, a customer first attitude, self starter with an entrepreneurial spirit, eager and ambitious"; "Curious, flexible and driven by innovation".

Job-seekers were informed that they would receive personalized feedback on their cover letter and CV, and that the three best cover letters (as evaluated by the firms' recruiters) would receive 500, 300, and 200 Euros respectively. Furthermore, they were informed that recruiters could ask to be put in touch with high-quality applicants. Anecdotal evidence suggests that students highly valued the personalized feedback, and that this was their main motivation for participating.[8]

Job-seekers were randomized across a control and treatment group. The control group was shown a 10-minute placebo guide on how to best use LinkedIn for the job search, and subsequently had 1 hour to write a cover letter. Students were informed that they would not have access to certain websites, including Large Language Models.[9] The treatment group were shown a 10-minute guide on how to improve prompt-writing on ChatGPT, before being asked to write a cover letter. Job-seekers in the treatment group were blocked from accessing the same websites as the control group, with the exception of Open AI's ChatGPT 3.5, which they were allowed to access through a free account provided by the researchers.[10] We tracked their browser history and ChatGPT conversation histories, which indicated that while 18% of applicants in the control group (unsuccessfully) tried to access ChatGPT, 95% of applicants in the treatment group used ChatGPT to write their cover letter. We initially pre-registered and planned to include a third experimental arm in which job-seekers received the same training as the control group but were also allowed to use ChatGPT for their cover letters. However, due to power concerns, we dropped this arm before the experiment began.

Randomization was stratified upon job-seeker's university, their gender, whether they were a Bachelor or Master student, their GPA, and age. Appendix Table A1 shows that randomization was successful, as groups are balanced across stratified variables. We observe that the ratings of CVs by the recruiters are higher in the Control than Treatment group (see Appendix Table A2), and hence control for the CV grade in our regressions, following Bruhn and McKenzie (2009). Since CVs were submitted before the experiment began, they could not have been influenced by the treatment.

Recruiters were blind to the two treatments and were told by their manager (who was our point of contact) that the purpose of the exercise was to test for the firm's internal

[8] The multiple incentives — including monetary incentives and the possibility of being hired — ensured participants submitted the best possible application, akin to a real hiring setting.

[9] See Appendix B.1.1 for a list of blocked websites.

[10] At the time of the experiment, Spring 2024, GPT 3.5 was the most advanced free version.

consistency across recruiters - and were thus not told who the other recruiter within the firm would be. The recruiters were further told to identify high-quality candidates that they would be interested in interviewing for (future) open positions.

To estimate the effect of the use of ChatGPT on the perceived quality of an applicant's cover letter and likelihood of getting interviewed, we run the following OLS specification:

$$Y_i = \beta_0 + \beta_1 ChatGPT_i + \gamma X_i + \mu_{f,r} + \mu_s + \varepsilon_i \tag{1}$$

where $ChatGPT_i$ is an indicator equal to one if the individual is assigned to the ChatGPT treatment, and 0 otherwise. X_i is a vector of baseline covariates that were unbalanced at baseline and were used to stratify randomization (Bruhn and McKenzie, 2009). We include firm-by-recruiter fixed effects ($\mu_{f,r}$) and school-level fixed effects (μ_s). Following Abadie et al. (2022), we cluster standard errors at the level of randomization, namely the job-seeker level since each applicant receives evaluations from multiple recruiters, creating potential correlation in the error terms across observations for the same individual.[11] Estimation is by OLS, except for when the outcome variable is *Likelihood of Interview* and *High Chance of Interview*, which are an ordered logit, and logit regression, respectively.

3 Job-seeker Experiment: Results

This section presents both pre-registered and exploratory results. The pre-registered outcome variables were the recruiter's evaluations, browser history, and LLM prompts, corresponding to Tables 1 and 4 and Figure 2. The remaining analyses were exploratory and not pre-registered.

3.1 Main Treatment Effects

Table 1 presents the main results of the experiment, showing the effect of ChatGPT usage on various aspects of the quality of the cover letter and the likelihood of being invited to a job interview, as evaluated by two independent recruiters. Outcome variables are standardized (except for the logit regressions), and therefore treatment effects are reported in terms of

[11] Table A12 reports treatment effects without clustering.

standard deviations. Table 1 illustrates that the use of ChatGPT has a positive impact on the overall quality of the cover letter written by an applicant, improving its quality by 0.222 standard deviations on average, which is statistically significant at the 10% level (column 1). This improvement is primarily driven by enhancements in the introduction and closing sections of the cover letter, with treatment effects of 0.253 and 0.281 standard deviations, respectively, both significant at the 5% level (columns 3 and 6). The use of ChatGPT does not have a statistically significant treatment effect on the cover letter's layout and clarity, nor the sections discussing one's experience or motivation.

Table 1: Main Results: Effect of ChatGPT on Cover Letter

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Cover Letter				Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
ChatGPT	0.222*	0.161	0.253**	0.075	0.150	0.281**	0.249	0.011
	(0.117)	(0.145)	(0.119)	(0.112)	(0.115)	(0.113)	(0.284)	(0.444)
Control Group Mean	0.000	0.000	0.000	0.000	0.000	0.000	2.799	0.278
Observations	274	274	274	274	274	274	274	274

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Standard errors are clustered at the individual level. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

While ChatGPT improves the evaluation of some aspects of the cover letters, it does not translate into a higher likelihood of being invited to the next phase of the recruitment process, a job interview. The causal estimate of the effect of LLM usage on a job-seeker's interview likelihood is positive but not statistically significant ($p = 0.380$, column 7).[12] Appendix Figure A2 presents the marginal treatment effects, showing that they are not only statistically insignificant but also economically negligible. Similarly, the probability of having a high chance of being invited to an interview (pre-registered as a *Likelihood of Interview* score greater than 3 on a 5-point scale) shows a marginally positive, but statistically and economically insignificant effect ($p = 0.981$, column 8). The results are robust to using the Belloni et al. (2014) Double LASSO approach (Appendix Table A6), and bootstrapped standard errors (Appendix Table A7), increasing our confidence that the estimated treatment

[12]The interpretation of the coefficient is different compared with column 1-6, as the regression is an ordered logit because the outcome variable is on a five-point scale.

effects are accurate despite the small sample size. This finding implies that the underlying matching between job-seekers and firms is unaffected by the usage of LLMs.

We compared ratings by ChatGPT to those of the recruiters, to assess whether recruiters themselves used LLMs to evaluate cover letters. Using the same evaluation criteria, ChatGPT graded the cover letters as well (see Appendix A.1.6). We find that ChatGPT, unaware of which letters it had assisted in writing, consistently assigned higher scores to those it helped produce. This discrepancy confirms that our recruiters did not rely on LLMs for their evaluations, and highlights a concerning bias in addition to those documented in the literature already (Hoffman et al., 2017; Avery et al., 2023), given the increased use of algorithmic screening tools in recruitment (Eric Reicin, 2021; Institute for the Future of Work, 2022).

To understand why improvements in the cover letter's quality do not translate into higher interview chances, we examine the relative importance of different aspects of the cover letter in determining the likelihood of being invited to a job interview. Table 2 shows that the cover letter's clarity and motivation section, are the most important factors in determining interview likelihood (significant at the 1-5% level). However, our results in Table 1 indicate that ChatGPT does not significantly improve either component, which can explain why ChatGPT enhances the overall quality of a cover letter, but does not increase the likelihood of being invited to an interview.

Table 2: Determinants of Interview Likelihood

	(1) Likelihood of Interview
CL: Layout	0.525**
	(0.216)
CL: Intro	0.293
	(0.249)
CL: Experience	0.399
	(0.249)
CL: Motivation	0.823***
	(0.249)
CL: Closing	0.474
	(0.315)
Control Group Mean	2.799
Observations	274

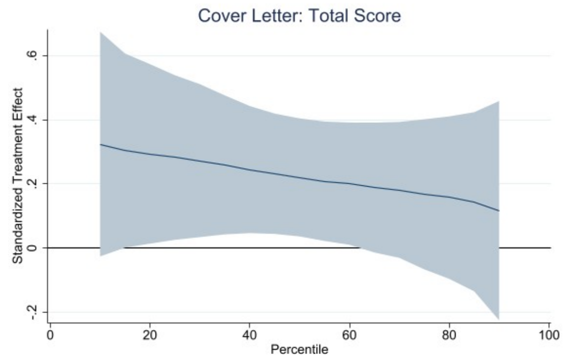
Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

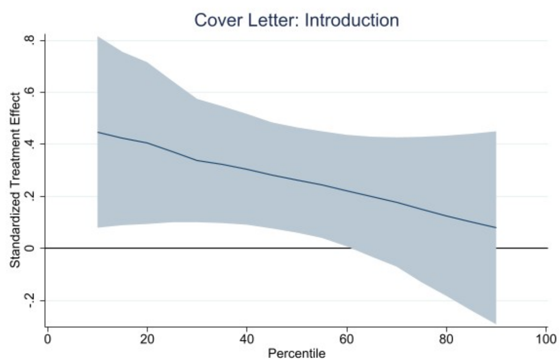
Notes: Ordered Logit regression, regressing the likelihood of an interview to a job interview, on the sub-components of the cover letter and CV (described in Appendix B.1.3). Sub-components of the CV are omitted from the regression table, for visibility purposes. Standard errors are clustered at the individual level. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Figure 1. Quantile Regressions

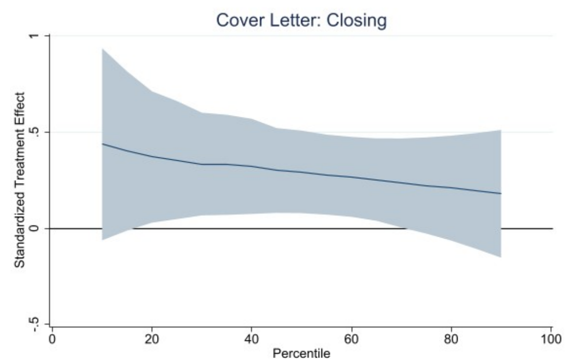
(a) Cover Letter: Total Score



(b) Cover Letter: Introduction



(c) Cover Letter: Closing



Notes: Each panel plots the estimated treatment effect of ChatGPT on cover letters across quantiles of the outcome distribution, obtained from quantile regressions. The outcomes correspond to the three main subcomponents of the cover letter—Total Score, Introduction, and Closing—as described in Appendix B.1.3. Shaded areas represent 95% confidence intervals based on robust standard errors clustered at the individual level. The corresponding average (OLS) treatment effects are reported in Table 1.

Decomposing the average treatment effects by the applicant's perceived ability, measured by the assigned grade on the cover letter, we find that the positive treatment effects of the use of LLMs on the cover letter quality are driven by lower-quality applicants, in line with the findings of Noy and Zhang (2023), Dell'Acqua et al. (2023), and Brynjolfsson et al. (2025).[13] Figure 1 presents the standardized treatment effects of quantile regressions,

[13]We do not find any heterogeneous treatment effects by gender of the job-seeker, or by type of university

illustrating a clear pattern of lower perceived ability applicants benefiting more from LLMs than their higher perceived ability counterparts. This trend is consistent across the overall cover letter score (Panel A), introduction (Panel B), and closing section (Panel C).[14]

For the total cover letter score (Panel A), the effect of ChatGPT is strongest at the lower end of the distribution, with a standardized treatment effect of 0.318 standard deviations at the 10th percentile. The introduction section (Panel B) shows a particularly pronounced decline in treatment effect across percentiles, from 0.433 at the 10th percentile to 0.076 standard deviations at the 90th percentile. The closing section (Panel C) exhibits a more gradual decline but maintains a positive treatment effect across most of the distribution.

3.2 ChatGPT's Impact on Writing Clarity

Our main results demonstrate that access to ChatGPT significantly improves recruiters' evaluations of cover letter quality, but does not have an effect on the perceived writing clarity of the cover letter, as evaluated by two independent recruiters. However, underlying this null result are two contrasting mechanisms. Following Wiles et al. (2025), we analyze objective measures of writing clarity to understand how access to ChatGPT transforms the underlying text along two dimensions: grammatical mistakes, and readability.

Table 3 presents regression results of the effect of access to ChatGPT on the grammar, readability, and length of the evaluated cover letters. Columns (1) and (2) illustrate that access to ChatGPT substantially reduce the number and rate of grammatical errors in the cover letter. The total number of errors and the error rate decrease by 37-39% as a result of access to ChatGPT, compared with the control group ($p < 0.01$). Appendix Table A14 illustrates that this improvement in the cover letter's technical quality is primarily driven by a reduction in spelling errors. Appendix Figures A3 and A4 show the full distribution and box plots of error rates across treatment and control groups.

Despite these clear improvements in technical quality, our analysis of readability metrics reveals a more complex picture. Column (3) of Table 3 show that the Flesch Reading Ease score — a readability measure assessing how easy a text is to understand, with lower scores indicating increased difficulty — of cover letters written with the assistance of LLMs

degree, as discussed in Appendices A.1.2 and A.1.3.

[14] The heterogeneous treatment effects for Layout, Motivation, and Experience are not statistically significant, see Appendix Figure A1.

Table 3: Effect of ChatGPT on Clarity and Readability

	(1)	(2)	(3)	(4)	(5)	(6)
	Grammatical Errors		Readability		Length	
	Total Errors	Error Rate	Flesch Reading Ease	Flesch-Kincaid Grade Level	Word Count	Average Word Length
ChatGPT	-4.358***	-0.011***	-4.451**	0.589	6.330	0.174***
	(1.341)	(0.003)	(1.849)	(0.373)	(22.966)	(0.062)
Control Group Mean	11.931	0.028	35.982	13.917	428.153	6.294
Observations	137	137	137	137	137	137

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Standard errors are clustered at the individual level. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Columns 1 and 2 refer to grammatical errors, identified using the R package 'languagetoolR '. Columns 3 and 4 report treatment effects of the Flesch Reading Ease, and the Flesch-Kincaid Grade Level, evaluated using the R package 'koRpus'. Columns 5 and 6 report treatment effects on the word count and average word length, evaluated using ChatGPT's API. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

decreases by 12%, compared with cover letters in the control group. This means that the cover letters have become *harder* and less accessible to read as a result of access to ChatGPT, despite the reduced error rate. This treatment effect is statistically significant at the 5% level. The Flesch Reading Ease score of the control group (35.982) indicates that the cover letter's complexity corresponds to a college-level or professional-level text, with LLM access further increasing the reading difficulty of the text. Relatedly, column (4) reports treatment effects on the Flesch-Kincaid Grade Level, an additional readability test that indicates the U.S. school grade level needed to understand a piece of text, with higher scores indicating a more complex text. While treatment effects of access to ChatGPT are marginally insignificant ($p = 0.11$), the coefficient's positive sign and magnitude imply that an additional 6 months of college-level education are required to understand the cover letter generated with the assistance of ChatGPT, compared to cover letters generated without ChatGPT. Appendix Figures A5 and A6 plot the full distribution and boxplots of readability scores across treatment groups.

To understand why the text complexity of the cover letters increases as a result of LLM access, despite reducing grammatical errors, we next look at columns (5) and (6) of Table 3. While the treatment does not have an effect on the length of the cover letter (column (5)), access to ChatGPT statistically significantly increases the average word length by 3% (column (6)). These findings suggest that ChatGPT used longer words, which in turn reduces the readability of the text.

This pattern of improved technical quality coupled with reduced readability stands in contrast to the findings of Wiles et al. (2025), who showed that algorithmic writing assistants (AWAs) improve both technical quality and readability simultaneously. This can be reconciled by understanding the differences between AWAs and LLMs: rather than simply cleaning up existing text like AWAs, LLMs are generative and hence can transform the writing in ways that, while technically more sophisticated, may actually reduce clarity and accessibility—as documented among the abstracts of scientific papers (Alsudais, 2025). This transformation involves substituting simpler language with more complex alternatives, longer sentences, and more sophisticated vocabulary. These findings can also explain why ChatGPT use does not alter the perceived quality in the *Layout, Writing Quality, and Clarity* subcomponent of the cover letter evaluation from the recruiters point of view: while the text has fewer grammar and spelling errors, it is now more complicated to read.

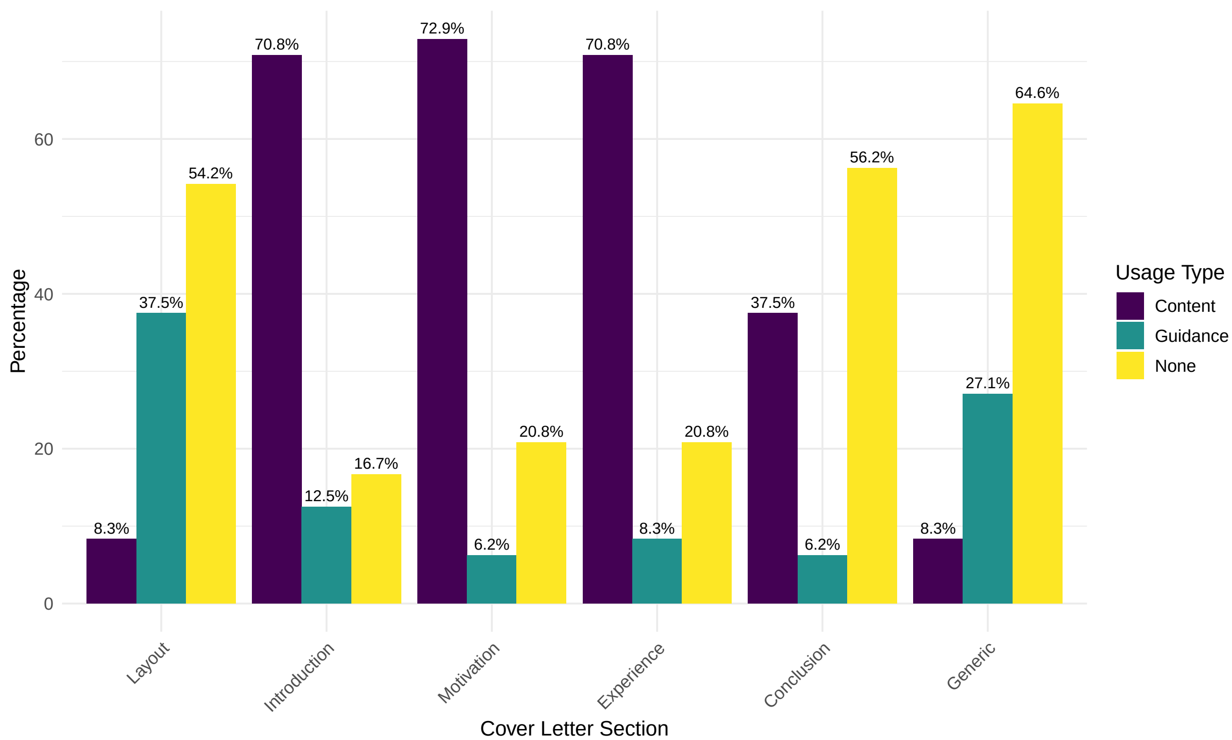
This difference has important implications for our understanding of how LLMs affect labor market signaling. While AWAs preserve and clarify the underlying informational content of cover letters and CVs, LLMs may obscure rather than reveal worker quality, contributing to the welfare losses documented in our structural model in Section 5. The fact that LLMs reduce the text's grammatical mistakes while reducing actual readability suggests that firms may find it increasingly difficult to extract meaningful information from LLM-assisted applications, even when those applications contain fewer grammatical mistakes.

3.3 Analysis of ChatGPT Interactions

In this sub-section, we explore how job-seekers interacted with ChatGPT, how access to ChatGPT affected their browsing history, and whether more effective prompting enhanced ChatGPT's effect.

We begin by analyzing the conversations between job-seekers and ChatGPT to understand how they engaged with the LLM during the cover letter writing process. We developed a comprehensive text analysis framework that quantifies both the focus and nature of user interactions. Our methodology tracks mentions of the six key cover letter sections (Layout, Introduction, Experience, Motivation, Conclusion, and Generic, which is added to the analysis to take into account requests that users make on the entire cover letter, rather than subsections), distinguishing between content-based interactions (where users seek substantive help with section content) and guidance-based interactions (where users ask for simple structural or formatting advice). For each conversation, we compute the percentage of total mentions per section and the ratio of content versus guidance requests. This allows us to understand both where users focus their attention and how they utilize ChatGPT. Appendix D provides the complete technical details of our methodology, with Appendix Table A34 and Figure A14 demonstrating substantial variation in user engagement, with the number of exchanges per conversation ranging from 1 to 16 messages (mean = 6.4, median = 5).

Figure 2. ChatGPT Prompts: Distribution of messages across cover letter sections.



Notes: The figure shows the distribution of ChatGPT usage across cover letter sections. Each bar reflects the share of conversations where the user requested Content (i.e., text generation or improvement), Guidance (i.e., structural or formatting advice), or showed No engagement with the section (None).

Figure 2 shows how users target their messages across different sections during the cover letter writing process, revealing that Introduction, Motivation, and Experience sections show the highest content engagement, with approximately 71%, 73%, and 71% content requests, respectively. In contrast, Layout and Generic sections receive notably different attention patterns, with Layout showing a low share on content-based interactions (8%) and higher in guidance (38%) requests, while Generic requests are also predominantly guidance-focused (8% content, 27% guidance). Therefore, the ineffectiveness of ChatGPT at improving the personalized sections of the cover letter is seemingly not due to job-seekers not asking for ChatGPT's advice on these sections. Figure 2 reveals that across most sections, users predominantly engaged with ChatGPT for content-based assistance rather than guidance, suggesting applicants use ChatGPT for substantial changes to the cover letter's content, rather than small changes in its format. The exceptions are the Layout and Generic sections, where users show greater interest in guidance rather than content-related assistance. This further underlines fundamental differences between LLMs and other technologies such as algorithmic writing assistants (Wiles et al., 2025).

Beyond section-specific engagement, we analyze four meta-categories of user behavior during cover letter writing. Figure A13 shows substantial engagement across these categories: 89% of users sought general strategy advice, 85% conducted company research, 60% made revision requests, and 17% focused on formatting issues. The high company research engagement (85%) indicates most job-seekers used ChatGPT as an information-gathering tool beyond writing assistance, seeking insights about prospective employers, industry trends, and organizational culture. This pattern suggests ChatGPT serves as a comprehensive research platform rather than merely a text generation tool. Given the extensive use of ChatGPT for company research and strategic planning, we expect to observe corresponding reductions in traditional online information-seeking behaviors.

Our next analysis therefore investigates whether ChatGPT complements or replaces other information tools. Table 4 reveals a significant substitution effect between ChatGPT usage and other online resources: job-seekers in the treatment arm conducted 51% fewer Google searches and visited 38% fewer websites compared to individuals in the control arm (columns 1 and 2). This effect is particularly pronounced for searches related to grammar, which decrease by 84% among individuals with access to ChatGPT (column 6). These findings suggest that ChatGPT was used as a comprehensive tool, potentially replacing the need for multiple online resources. The marked reduction in grammar-related searches indicates that users may rely on ChatGPT's language proficiency, reducing their dependence on external grammar tools, such as those evaluated in Wiles et al. (2025). This substitution effect highlights ChatGPT's capacity to streamline the writing process by consolidating various aspects of language assistance and information gathering into a single platform.

Browser histories are not available for all job-seekers, with differential availability across experimental arms (see Appendix Table A9).[15] Appendix Table A11 presents Lee (2009) bounds to account for potential selection bias, confirming the robustness of our findings. The consistent negative treatment effects on websites visited across these bounds further support the conclusion that ChatGPT usage substantially alters online information-seeking behavior during the cover letter writing process.

Table 4: Browser History

	(1)	(2)	(3)	(4)	(5)	(6)
	# Websites Visited	# Google Searches	Searched: Firm	Searched: Cover Letter	Searched: Grammar	Searched: Translation
ChatGPT	-7.422***	-5.120***	-1.825	-1.140	-2.427***	-1.280
	(2.215)	(1.320)	(1.169)	(0.692)	(0.631)	(0.749)
Control Group Mean	19.351	10.000	5.281	2.368	2.895	2.175
Observations	119	119	119	119	119	119

Notes: Intention to Treat estimates. Column 1 reports treatment effects on the total number of websites visited, while Column 2 refers to the number of google searches. Columns 3-6 refer to whether the subject searched the relevant topic. Treatment effects are reported from OLS regressions. Standard errors are clustered at the individual level. PD Lasso machine learning technique is used to select control variables (Belloni et al., 2014), along with firm and school-level fixed effects. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. The sample consists of job-seekers in the Control and Treatment arms, for whom browser histories were saved. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Lastly, we examine whether adherence to the ChatGPT training influenced cover letter quality, aiming to disentangle the effect of LLM usage from the guidance job-seekers received. To assess the impact of training adherence, we use OpenAI's API to perform sentiment analysis on job-seekers' conversation histories with ChatGPT, measuring the extent to which they applied the suggested prompting techniques. Table 5 reports the normalized treatment effect of compliance with the ChatGPT training on cover letter quality. The results show no significant relationship between training adherence and cover letter scores. This suggests two possibilities: either improved prompting does not meaningfully enhance cover letter quality in this context, or our training was insufficiently effective in teaching techniques that translate into better outcomes.[16] These precisely estimated null results reinforce our confidence that the observed treatment effects are primarily driven by ChatGPT usage rather than the training itself. The pre-registered study design included an additional treatment arm to causally disentangle whether the treatment effect is driven by the prompting guide or access to ChatGPT, however this treatment arm was dropped due to power concerns. The correlational evidence presented in Table 5 suggests that the prompting guide did not amplify

[15] When collecting browser history and ChatGPT conversations, 18 computers did not contain any recorded history.

[16] The training's content is similar to that recommended by OpenAI (see here), despite being developed independently, 10 months earlier.

the effectiveness of the ChatGPT intervention, in line with recent studies (Bashardoust et al., 2024). Relatedly, Appendix Table A13 reports the correlational relationship between cover letter quality and number of ChatGPT prompts, with no persistent relationship found.

Table 5: Effect of ChatGPT Training on Results

	(1)	(2)	(3)	(4)	(5)	(6)
	Cover Letter					
	Total	Layout	Intro	Experience	Motivation	Closing
ChatGPT Training Compliance	-0.001	0.021	0.046	-0.111	-0.010	0.046
	(0.100)	(0.115)	(0.084)	(0.070)	(0.099)	(0.107)
Control Group Mean	-0.056	-0.051	-0.090	0.009	-0.021	-0.094
Observations	82	82	82	82	82	82

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Standard errors are clustered at the individual level. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Columns 1-6 refer to variables as described in Appendix B.1.3. ChatGPT training compliance is a 10-point scale indicating whether individuals applied the ten techniques discussed for effective prompting, see Appendix B.1.2. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively. The sample consists of job-seekers in the Treatment arm, for whom ChatGPT conversation histories were saved.

4 Recruiter Experiment: Perceptions of AI

The first experiment reveals that while job applicants benefit from using LLMs in their cover letters, the improvements are limited to the less critical sections. Furthermore, although access to LLMs reduced the number of spelling and grammar mistakes, it also reduced the cover letter's readability. As a result, these enhancements do not significantly impact the likelihood of securing a job interview. However, it remains unclear if recruiters can detect LLM-generated cover letters and what their attitudes towards LLM usage are. Strong preferences could bias hiring decisions, leading to outcomes based on the recruiters' perceptions rather than the quality of the job-seeker's signal. This raises an important question: what happens to recruiters' evaluations of cover letters if they are explicitly informed that some applicants used LLMs? To address this, we run a second experiment with 401 recruiters on Prolific, examining the role of LLM disclosure and its potential influence on recruiters' decisions.[17]

4.1 Experimental Design

The recruiters are assigned to evaluate five pseudonymized cover letters, and are asked to evaluate each cover letter using the same grading rubric as recruiters in the job-seeker experiment.[18] The cover letters are drawn from the sample of cover letters that were written by job-seekers in the job-seeker experiment.[19] 84% of recruiters are employed either full- or part-time, 59% identify as female, with an average age of 37.[20] They are randomized across three treatments: *No Information*, *Partial Information*, and *Full Information*.

In the *No Information* treatment, recruiters are asked to evaluate the quality of the cover letters without receiving further information about the nature of the job-seeker experiment. This captures the naive setup, where recruiters are unaware of the two experimental arms, and hence are in the same situation as the recruiters in the job-seeker experiment (and real world). In *Partial Information*, recruiters are told that some cover letters were written with the assistance of ChatGPT, while others were not. In addition to evaluating the quality of the cover letters, recruiters in this treatment are asked to identify which cover letters are written with ChatGPT's assistance. In *Full Information*, recruiters are informed precisely which cover letters are written with the assistance of ChatGPT. This explores the scenario where applicants have to disclose LLM use in their application (which 49% of recruiters are in favor of), or if humans were able to diagnose the use of LLMs, and hence how recruiters respond to the use of LLMs by job-seekers.

[17] Among the recruiters, 65.3% are from the United Kingdom, while the remaining 34.7% come from various European countries: Germany, Italy, Portugal, Spain, France, the Netherlands, Ireland, the Czech Republic, Denmark, Finland, Switzerland, Belgium, Austria, and Sweden. All recruiters work in Human Resources, are fluent in English, and have been active on Prolific within the past 90 days.

[18] Recruiters are paid a fixed fee. This is in line with Carvajal et al. (2024), and several other recent studies that use unincentivized measures (Ameriks et al., 2020; Andre et al., 2022; Stango and Zinman, 2023), motivated by the literature finding limited differences between real-life behavior and related unincentivized measures (Brañas-Garza et al., 2021, 2023; Falk et al., 2023).

[19] Six cover letters are selected from both the control and treatment group in the job-seeker experiment. Cover letters are chosen such that for both treatment and control, two cover letters are from the bottom, middle, and top tercile of the total grade distribution. Furthermore, an even gender split was ensured, and all cover letters applied to the same job description. The average grades of the cover letters, as evaluated by the recruiters in the job-seeker experiment, are 5.90 vs. 5.85 for control and treatment groups ($p = 0.95$).

[20] A recruiter in our study had participated in 540 other studies on Prolific, took 20 minutes to complete the survey, and received an hourly payment of \$13.21, on average.

We run the following OLS specification to estimate the treatment effect of *Partial Information* and *Full Information* on the recruiter's evaluation of the applicant's cover letter and likelihood of getting interviewed, for cover letters written with and without the assistance of LLMs:

$$\begin{aligned} Y_{a,r} = \beta_0 &+ \beta_1 PartialInfo_r + \beta_2 FullInfo_r + \beta_3 CoverLetterGPTwritten_r \\ &+ \beta_4 PartialInfo * GPTwritten_r + \beta_5 FullInfo * GPTwritten_r \\ &+ \gamma X_r + \mu_a + \varepsilon_{a,r} \end{aligned} \tag{2}$$

where $Y_{a,r}$ is the outcome variable for the cover letter written by job applicant a, evaluated by recruiter r. $PartialInfo_r$ and $FullInfo_r$ are indicators equal to one if the recruiter is assigned to the *Partial Information* and *Full Information* treatment, respectively, and 0 if the recruiter is assigned to the *No Information* treatment. $CoverLetterGPTwritten_r$ is an indicator variable if the cover letter was written by a job-seeker who had access to LLMs, and zero otherwise. We include interactions terms between treatment assignment and the nature of the cover letter, and we control for the recruiter's age and sex (X_r) and job applicant-level fixed effects (μ_a). Following Abadie et al. (2022), we cluster standard errors at the recruiter level since each recruiter evaluates multiple cover letters, creating potential correlation in the error terms across observations from the same recruiter.[21] Estimation is by OLS, except for when the outcome variable is *Likelihood of Interview* and *High Chance of Interview*, which are an ordered logit, and logit regression, respectively. All of the results presented in section 4.2 were pre-registered.

4.2 Results

Table 6 reports average treatment effects of the *Partial Information* and *Full Information* treatments on the quality of the cover letter and likelihood of inviting the candidate to an interview, for cover letters written with and without the assistance of ChatGPT. By comparing *No Information* and *Partial Information*, we observe whether priming recruiters on the potential use of ChatGPT influences their perception of the cover letter's quality. In line with the fact that recruiters already believed that 61% of applicants use LLMs in their

[21] Table A27 reports treatment effects without clustering.

job applications, we find no statistically significant differences between *No Information* and *Partial Information*. Furthermore, in only 49% of cases in the *Partial Information* treatment did recruiters correctly identify whether the cover letter was written with the assistance of LLMs, statistically indistinguishable from guessing ($p = 0.5114$).[22] This provides evidence that while recruiters are aware that applicants use LLMs in their job applications, they are unable to identify which applicant actually uses LLMs.[23] We furthermore do not observe a statistically significant difference in the *No Information* treatment between the evaluation of cover letters written with and without ChatGPT.

The *Full Information* treatment does not result in a higher evaluation of the overall grade of the cover letter (column 1), both for cover letters written with and without access to LLMs. Despite no difference in the evaluation of the overall quality of the cover letters, recruiters are more likely to invite job-seekers to the next stage of the recruitment process in the *Full Information* treatment when the cover letter is written without the assistance of LLMs, as indicated by the statistically significant increase in the high chance of inviting the candidate to an interview (column 8).

Comparing the *Partial Information* and *Full Information* treatments, an interesting pattern emerges: while evaluations of cover letters written with ChatGPT are unchanged between the two treatments, recruiters in the *Full Information* treatment evaluated cover letters written without access to ChatGPT (which the recruiters in that treatment arm knew with certainty) statistically significantly more positively, compared to recruiters in the *Partial Information* treatment. All components of the cover letter written without access to ChatGPT — aside from the Motivation section — are evaluated more positively ($p = 0.03 - 0.08$), resulting in a statistically significantly greater likelihood of inviting the job-seeker to the next stage of the recruitment process. This indicates that recruiters are not penalizing applicants for using LLMs, but instead reward applicants that do not use them.[24]

Taking advantage of our experimental design, we can gain further insights into what

[22] The ability to correctly detect LLM usage is uncorrelated with the recruiter's confidence at being able to detect LLM usage ($\rho = 0.0292, p = 0.7409$).

[23] Young recruiters (below the median age) are more confident in their ability to detect LLM usage than older applicants ($p = 0.0405$), however we detect no statistically significant difference in their actual ability to detect LLM usage ($p = 0.8700$). Male and female recruiters are statistically indistinguishable in both their confidence and ability to detect LLM usage ($p = 0.5106$ and $p = 0.5829$, respectively).

[24] Appendix A.2.2 decomposes treatment effects by the gender of the recruiter, as well as their age. The positive treatment effects are primarily driven by female and older recruiters.

Table 6: Main Results: Effect of Knowledge About ChatGPT Usage on Cover Letter Evaluation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter			Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
Partial	-0.03	-0.10	-0.01	-0.02	-0.04	-0.05	-0.14	-0.16
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.15)	(0.16)
Full	0.06	0.08	0.12	0.10	0.09	0.10	0.19	0.26*
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.16)	(0.15)
Cover Letter Written with GPT	-0.04	0.03	0.01	0.00	-0.01	-0.06	-0.10	-0.09
	(0.07)	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)	(0.14)	(0.14)
Partial × Written with GPT	0.04	0.14	0.03	0.02	0.05	0.06	0.14	0.14
	(0.10)	(0.10)	(0.10)	(0.09)	(0.10)	(0.10)	(0.18)	(0.21)
Full × Written with GPT	0.02	0.02	-0.04	0.02	-0.07	-0.01	-0.15	-0.25
	(0.10)	(0.10)	(0.09)	(0.09)	(0.10)	(0.09)	(0.19)	(0.20)
t-test Partial: ChatGPT vs. No	0.99	0.02	0.64	0.78	0.57	0.99	0.61	0.76
t-test Full: ChatGPT vs. No	0.82	0.45	0.71	0.77	0.29	0.34	0.12	0.03
t-test Partial vs. Full: No ChatGPT	0.12	0.03	0.08	0.03	0.19	0.03	0.08	0.03
t-test Partial vs. Full: Yes ChatGPT	0.31	0.41	0.40	0.11	0.92	0.28	0.82	0.80
Control Group Mean	0.00	0.00	0.00	0.00	0.00	0.00	3.40	0.53
Control Group S.D.	1.00	1.00	1.00	1.00	1.00	1.00	1.20	0.50
N	2005	2005	2005	2005	2005	2005	2005	2005

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses, and are clustered at the recruiter-level. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

sort of cover letters are evaluated differentially when recruiters are made aware of LLM usage. The cover letters selected from the job-seeker experiment were balanced across LLM usage vs. not, but also balanced in terms of their quality. For cover letters written both with and without the assistance of ChatGPT, two were chosen from the bottom, middle, and upper tercile, respectively. Hence we have comparable heterogeneity in the quality of cover letters. Appendix A.2.5 decomposes the average treatment effects from Table 6 for cover letters that scored in the lower and middle tercile. While cover letters in the bottom tercile written with LLM assistance scored statistically significantly better, there were no statistically significant differences across the three treatment arms. For cover letters ranked in the medium tercile, there is also no statistically significant effect of revealed LLM usage on the recruiters' evaluation of the quality of the cover letter. Nevertheless, recruiters in the *Full Information* treatment are more likely to invite applicants with average-quality cover letters to an interview when informed that the cover letter was written without LLM assistance. This suggests that awareness of LLM usage influences recruitment decisions, specifically whether an applicant is invited for an interview. While the use of LLM itself does not directly affect hiring decisions or labor market outcomes, knowledge of its use may have negative consequences for job-seekers.

Table 7 reports treatment effects on high quality cover letters, as identified by recruiters in the job-seeker experiment. In the *No Information* treatment, recruiters evaluate cover letters written with ChatGPT assistance statistically significantly worse, also reducing the likelihood of inviting the job-seeker to an interview. This pattern is replicated in the *Partial Information* treatment, where cover letters written with ChatGPT are also evaluated statistically significantly worse. However, this does not translate into a lower likelihood of being interviewed ($p = 0.14$). In the *Full Information* treatment, the difference between cover letters written with and without LLM assistance — which the recruiters knew — was the most pronounced, with highly statistically significant treatment effects of the cover letter being written with LLM assistance on the overall rating of the cover letters, its individual components, and the likelihood of being invited to a job interview ($p = 0.00 - 0.05$). Therefore, recruiters reward, high-quality cover letters when they are certain that they are written independently, without LLM assistance.

Comparing the *Partial Information* to the *Full Information* treatment highlights interesting conclusions too: while the recruiter's evaluations for cover letters written with the assistance of ChatGPT do not differ statistically significantly between the two treatments, the evaluations of cover letters written without LLM assistance do. Compared to the *Partial Information* treatment, recruiters in the *Full Information* treatment arm evaluated cover letters written without LLM assistance significantly more positively when informed that the (high-quality) cover letter was written without the assistance of ChatGPT. All dimensions of the cover letter are evaluated more positively (columns 2-6), which subsequently translates into large and highly significant increases in the likelihood of inviting the candidate to a job interview (columns 7-8).

In summary, we observe that recruiters are unable to correctly identify the use of LLMs in job applications: they perform no better than chance. However, when recruiters are informed whether job-seekers used LLMs, their evaluations of the quality of the cover letter, and likelihood to invite the applicant to a job interview, adjust. Disclosing LLM usage has no effect on the perceived quality of low- and medium-quality cover letters, but recruiters evaluate high-quality cover letters more positively when these are not written with

Table 7: Main Results: Effect of Knowledge About ChatGPT Usage on Cover Letter Evaluation - Upper Tercile

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter			Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
Partial	-0.16	-0.25**	-0.05	-0.08	-0.20*	-0.22*	-0.42*	-0.56*
	(0.12)	(0.13)	(0.13)	(0.12)	(0.12)	(0.11)	(0.25)	(0.29)
Full	0.09	0.11	0.18	0.13	0.06	0.05	0.19	0.26
	(0.11)	(0.12)	(0.12)	(0.11)	(0.12)	(0.11)	(0.24)	(0.31)
Cover Letter Written with GPT	-0.48***	-0.29**	-0.22	-0.28**	-0.48***	-0.69***	-0.92***	-0.90***
	(0.12)	(0.13)	(0.13)	(0.12)	(0.12)	(0.11)	(0.27)	(0.30)
Partial × Written with GPT	0.20	0.29	0.05	0.03	0.31*	0.34**	0.51	0.54
	(0.17)	(0.18)	(0.19)	(0.16)	(0.17)	(0.17)	(0.36)	(0.40)
Full × Written with GPT	0.04	0.05	-0.12	-0.03	0.12	0.11	-0.03	-0.08
	(0.16)	(0.17)	(0.18)	(0.17)	(0.17)	(0.16)	(0.35)	(0.42)
t-test Partial: ChatGPT vs. No	0.03	0.99	0.22	0.03	0.19	0.01	0.14	0.19
t-test Full: ChatGPT vs. No	0.00	0.05	0.01	0.01	0.00	0.00	0.00	0.00
t-test Partial vs. Full: No ChatGPT	0.05	0.00	0.11	0.02	0.06	0.10	0.08	0.01
t-test Partial vs. Full: Yes ChatGPT	0.53	0.29	0.71	0.18	0.62	0.85	0.98	0.47
Control Group Mean	0.18	0.18	0.13	0.20	0.12	0.10	3.66	0.62
Control Group S.D.	0.89	0.93	0.99	0.93	0.95	0.94	1.08	0.49
N	668	668	668	668	668	668	668	668

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses, and are clustered at the recruiter-level. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Sample consists of cover letters that were identified as being part of the upper tercile, based on evaluations from the first Experiment. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

the assistance of LLMs. This sizable increase in perceived quality also translates into a far greater likelihood of inviting the job applicant to an interview.

5 Assignment Model with Imperfect Information and LLMs

Our experiments highlight the effects of LLMs at a specific moment and under an exogenous adoption rate of this new technology. To better understand potential efficiency losses arising from the use of LLMs during the application process, we run a calibration exercise. Building on the literature on labor market signaling (Spence, 1973; Kurlat and Scheuer, 2020; Bassi and Nansamba, 2021; Carranza et al., 2022) and matching (Teulings, 1995; Eeckhout and Kircher, 2018), we apply the introduction of LLMs to a static one-to-one assignment model with imperfect information, where firms cannot directly observe workers' true abilities and instead rely on noisy signals.[25] Based on our findings from the first experiment, we introduce LLMs as a new signal-enhancing technology, which disproportionately improves the signals of lower-ability workers, creating a non-linear shift in the distribution of signals (cover letter) received by firms. Firms form Bayesian estimates of worker abilities as well as probability that LLMs were employed based on these signals, and matches are determined by positive assortative matching based on these estimates. Our exercise shows that while LLMs may improve signals for some workers, the resulting non-linear distortion introduces additional uncertainty for the firms, ultimately reducing aggregate matching efficiency relative to scenarios with either perfect information or symmetric noise without LLMs.

We focus on positive assortative matching (PAM) because it best captures the structure of the labor markets relevant to our empirical setting: university graduates applying for entry-level, full-time positions. PAM implies that higher-ability workers tend to match with higher-quality firms, leading to a sorting pattern where talent is concentrated in more selective jobs, while lower-ability workers sort into less demanding roles (Arcidiacono et al., 2010). This assumption is particularly plausible for the industries represented in our first experiment: large multinational firms operating in the banking, insurance, health, and technology sectors, which regularly attract the best graduates from selective STEM-oriented universities

We therefore develop a model that generalizes this empirical setting. In particular, the heterogeneous effects we observe in our first experiment — where LLMs disproportionately inflate signals for lower-ability job-seekers — interact meaningfully with PAM. In a positively assortative labor market, signal distortion can compromise efficient sorting by obscuring the true ranking of candidates, leading to misallocation of talent. Low-ability applicants who

[25] Wiles et al. (2025) also develop a model of firms hiring workers who are unable to observe the workers' productivity and hence rely on noisy signals. Nevertheless, our models differ on several key dimensions. Firstly, in Wiles et al. (2025), there are two types of workers, H and L, where only H type workers benefit from AWAs by reducing the effort cost of creating a high-quality signal or removing their "bad writing" shock. In contrast, in our model LLMs, in line with findings from Experiment #1, disproportionately benefit lower-quality workers. Therefore, in our model LLMs decrease the dispersion of signals, while AWAs increase it in Wiles et al. (2025).Secondly, in Wiles et al. (2025), firms are perfect substitutes and pay workers their expected productivity. Instead, our model is characterized by heterogeneity in both firms and workers and positive assortative matching, and hence more productive firms want to hire the most productive workers. As such in our two-sided model we find that LLMs reduce total output by causing more mismatches between workers and firms, while the partial equilibrium model of Wiles et al. (2025) has an ambiguous effect on welfare as high-quality workers gain at the cost of low-quality ones.

produce inflated application materials may displace more suitable candidates at high-quality firms, while those firms, unable to discern signal quality, may reduce their reliance on signals sent by applicants altogether.

The model allows us to formally evaluate this interaction by constructing a counterfactual scenario in which LLMs are not adopted. Comparing equilibrium outcomes across these two settings enables us to assess how LLM-induced distortions affect the allocation of talent across firms. We calibrate the model, disciplining the non-linearity in the distortion based on our estimates from the first experiment, and the moments relative to the job-seekers and firms' distribution from the literature (Teulings, 1995; Eeckhout and Kircher, 2018). As the calibration shows, welfare losses relative to the no-LLM benchmark can reach up to 0.8%, highlighting the potential for significant misallocation in high-stakes labor markets.

An analogous distortion appears in the education literature on grade inflation. In signaling models such as Chan et al. (2007), grade inflation benefits lower-ability students more than high-ability ones. Although there are individual winners and losers, the overall talent allocation typically deteriorates. In our setting, LLMs produce a comparable improvement to labor market signals by disproportionately helping lower-quality applicants. As such, signals in the labor market become less informative, undermining efficient matching. Closest to our setting is Schwager (2012), who embeds grade inflation into a one-to-one assignment model with imperfect information and shows that low-ability students gain most while aggregate output falls relative to a noisy — but undistorted — grading benchmark. We reach a similar conclusion: under imperfect information, the ability-dependent, non-linear distortions introduced by LLMs reduce aggregate welfare relative to a benchmark without LLM adoption.

Assignment model. The static economy consists of a continuum of workers, indexed by their quality s and distributed with CDF G_s, and a continuum of firms, indexed by their quality x and distributed with CDF G_x. We assume that both G_s and G_x admit densities denoted by g_s and g_x and have positive and bounded supports $\mathcal{S} := [\underline{s}, \overline{s}]$ and $\mathcal{X} := [\underline{x}, \overline{x}]$. The assumption on bounded supports is made for simplicity and does not change our proofs. The value of a match between a worker of type s and a firm of type x is given by $f(x, s)$ which satisfies Assumption 1:

Assumption 1. We assume that f is increasing, concave in both arguments, and features complementarity:[26]

$$f(x', s') - f(x, s') \geq f(x', s) - f(x, s) \; \forall s' \geq s \; \forall x' \geq x \tag{3}$$

Equation (3) means that an incremental gain in the value of a match to having a higher x (x' instead of x) is higher when s is higher (s' instead of s). The reverse is also true: having a higher s increases the value of a match when x is also higher. As such, our model implies a degree of complementarity between firms and workers, meaning that the best firms will be matched with the best workers, and the worst firms with the worst workers. These properties treat f as a standard production function akin to the specifications in Schwager (2012) and Eeckhout and Kircher (2018).

We assume that firms do not observe workers' true abilities. Instead, each worker sends a noisy signal of her ability to all firms. After observing these signals, the firms form a positive assortative matching based on the estimated worker qualities, denoted by $\tilde{s}$. The assignment in this economy is given by a function $x = \sigma(\tilde{s})$, which matches each worker s to a firm x according to the firm's estimate $\tilde{s}$. In equilibrium, all firms and workers must be matched one-to-one.[27] Hence, the assignment of a firm x to an estimated worker quality $\tilde{s}$ is deterministic and expressed by $x = \sigma(\tilde{s}) = G_x^{-1}(G_{\tilde{s}}(\tilde{s}))$ where $G_{\tilde{s}}$ is the distribution of estimated qualities of all workers. All randomness in matching then stems purely from the noise in the workers' signal, which causes the estimated $\tilde{s}$ to deviate from the true worker quality s. In other words, conditional on the specific signal generated by a job-seeker, their matching to a particular firm is deterministic. Intuitively, this corresponds to a setting where each worker sends a single application (containing their noisy signal) to all firms, and neither the application nor vacancy-posting process generates additional costs in our model.

We make the following assumptions about the signals. A fraction p of workers sends a noisy linear signal, with the noise having zero mean and being uncorrelated with the worker's true ability. This fraction represents the portion of the population who either do not have access to, or choose not to adopt, LLM technology. In contrast, a fraction $1 - p$ of the

[26] This definition of complementarity is also known as weak supermodularity, and is equivalent to a non-negative cross derivative when it exists: $f_{xs} \geq 0$.

[27] Consequently, σ must be *measure-preserving*, ensuring an equal mass of workers and firms. See Appendix E for further details on the derivation of the assignment rule.

population adopts a technology (LLMs) that improves the quality of their signals.[28] These workers send a nonlinear signal that is increasing in their true ability. Consider a worker of true quality s. With probability p, the worker does not use LLM technology to enhance the signal. The signal y observed by firms is thus given by

$$y = \begin{cases} y_1 = s + e & \text{with probability } p, \\ y_2 = h(s + e) & \text{with probability } 1 - p. \end{cases} \tag{4}$$

where y_1 is the signal produced by the worker with quality s without the use of LLMs, and e is a mean-zero random variable. On the other hand, y_2 is the signal produced by the worker with quality s who uses LLM in producing the labor market signal. In our first experiment with job-seekers, we find that LLMs primarily improve the signals of low-ability workers while leaving high-ability workers' signals essentially unaffected. Motivated by this, we impose the following assumption on the function h.

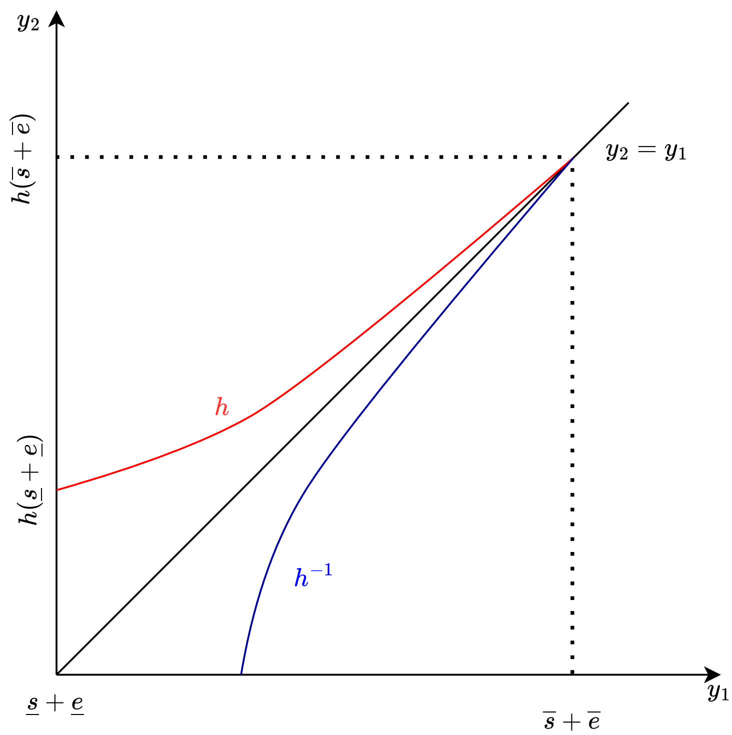
Assumption 2. $y_2 = h(y_1)$ is a continuous function that satisfies

1. $h(y_1) \geq y_1$ for all y_1
2. $0 < h'(y_1) < 1$ for all y_1
3. $\overline{y} := h(\overline{s} + \overline{e}) = \overline{s} + \overline{e}$

Assumption 2 implies that the technology used to increase the value of signals improves low-quality signals by a higher amount than high-quality ones, as we observed in the job-seeker experiment. An illustration h is shown in Figure 3.

[28]For simplicity we assume that the probability of adopting the technology is independent of s and x. In Appendix E we present an extension of the model where LLM usage is correlated with worker quality s.

Figure 3. Illustration function h



Notes: The function $h(y)$ captures the non-linear effect of LLMs on the signal y,with larger improvements for low-quality signals and smaller gains for high-quality ones.

An important feature of this information structure is that firms only observe signal y from workers without having knowledge on whether the signal is augmented by LLMs or not. This is supported by our second experiment, where we find that recruiters cannot properly identify whether a job application package was written with LLM assistance or not.

Our transformation function h echoes the "grade-inflation" mapping in Schwager (2012) as it affects job-seekers non-linearly relative to their type, so the incremental gain in the observed signal is largest for applicants whose original ability is lowest. This contrasts with the clarity view in Wiles et al. (2025). In their experiment, AWAs act as noise reducers: they correct spelling and grammar for every applicant, leaving the relative ranking of signals—and therefore the correspondence between the observed text and underlying ability—largely intact. Firms thus obtain a sharper, more precise signal of each candidate's true quality without altering the ranking of applicants.

Our experimental evidence shows that LLMs affect labor market signals quite differently. Because the quality boost diminishes as baseline ability increases, h compresses the signal distribution: low-ability applicants improve substantially while high-ability applicants benefit only marginally. Recruiters, unable to discern whether an observed signal is authentic or LLM-augmented, face greater uncertainty about the applicant's true ability. Capturing this non-linear distortion is therefore essential, as it affects the overall matching between firms and workers, and ultimately aggregate output. It moves the analysis beyond the clarity framework, highlights an uneven signal-amplification channel absent in prior work, and allows us to assess its equilibrium consequences for sorting and welfare.

From the signal y, firms form a Bayesian estimate of the worker's ability $\tilde{s}(y) = \mathbb{E}[s|Y = y]$ without ex ante knowing whether the signal was generated by LLMs or not. But firms can estimate the probability of the usage of LLMs based on the signal they receive. Using the law of total expectations, this estimate of worker's ability conditional on signal y can be decomposed as:

$$\tilde{s}(y) = \omega(y)\mathbb{E}[s|Y_1 = y] + (1 - \omega(y))\mathbb{E}[s|Y_2 = y] \tag{5}$$

where $\omega(y)$ is the posterior probability of LLM being used conditional on signal y, i.e. the firm's best estimate of whether or not LLM was used in the cover letter. Using Bayes' rule this is given by the likelihood ratio (where g_z denotes PDF of random variable z).

$$\omega(y) = p\frac{g_{y_1}(y)}{g_y(y)} = p\frac{g_{y_1}(y)}{pg_{y_1}(y) + (1 - p)g_{y_2}(y)} \tag{6}$$

If h satisfies Assumption 2, as shown in figure 3, this means that the LLM branch of the signal, Y_2, will take values of at least $\underline{y}_2 := h(\underline{s} + \underline{e})$. So for signals $y < \underline{y}_2$, $g_{y_2}(y) = 0$ and thus $\omega(y) = 1$ and $\tilde{s}(y)$ collapses to $\tilde{s}(y) = \mathbb{E}[s|Y_1 = y]$. In other words, if the signal quality is too low, the firm can know with certainty that the signal was not generated with the assistance of LLMs (see Figure 3). For all other signal values, the law of total expectation is used to decompose firms' estimates on the job-seekers' quality.

We denote the first expectation in Equation (5) by $\hat{s}(y) := \mathbb{E}[s|Y_1 = y]$. This expectation in fact coincides with the estimate of a workers' ability in the absence of LLMs, i.e. when LLMs are not used by any job applicant and the firms know this with certainty. Since h is monotonic and thus invertible, the second conditional expectation is equivalent to

$\mathbb{E}[s|Y_1 = h^{-1}(y)] = \hat{s}(h^{-1}(y))$.[29] Therefore, the estimate of workers' ability can be written as:

$$\tilde{s}(y) = \begin{cases} \hat{s}(y) & , \text{ for } y \leq \underline{y}_2 \\ \omega(y)\hat{s}(y) + (1-\omega(y))\hat{s}(h^{-1}(y)) & , \text{ for } y \geq \underline{y}_2 \end{cases} \tag{7}$$

Since $h^{-1}(y) \leq y$ we have the following inequality

$$\tilde{s}(y) < \hat{s}(y), \quad \underline{y}_2 \leq y < \overline{y} \tag{8}$$

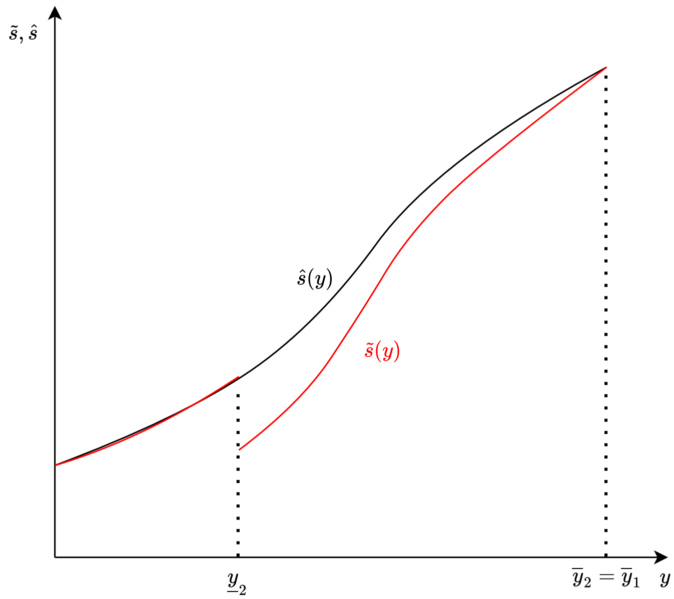
which, as illustrated in Figure 4, shows that for the range of signal values where the signal could have been augmented by LLMs ($\underline{y}_2 \leq y < \overline{y}$), firms form a lower estimate of workers' ability than the case where *all* signals are generated without LLM assistance.

For the range of signal values where firms cannot ascertain whether the signal was augmented by LLMs, they form a universally lower estimate of worker ability compared to the case without LLM assistance. This occurs because firms' Bayesian estimates incorporate the possibility of augmentation, effectively discounting the observed signal, leading to less informative signals. Consequently, while LLMs improve the quality of lower signals, they introduce a new form of uncertainty, which can negatively affect firms' perceptions of worker ability within this range.

[29] To see this, consider $\mathbb{E}[s|Y_2 = y] := \int s g_{s|y_2}(s|y)\mathrm{d}s$, and the CDF of y_2 can be written as $G_{y_2}(x) = \Pr\{Y_2 \leq x\} = \Pr\{h(s+e) \leq x\} = \Pr\left\{y_1 \leq h^{-1}(x)\right\} = G_{y_1}(h^{-1}(x))$. Similarly $G_{y_2|s}(x|s) = G_{y_1}(h^{-1}(x)|s)$. Then by Bayes' rule:

$$g_{s|y_2}(s|y) = \frac{g_{y_2|s}(y|s)g_s(s)}{g_y} = \frac{g_{y_1|s}(h^{-1}(y)|s)g_s(s)}{g_{y_1}(h^{-1}(y))} \equiv g_{s|y_1}(s|h^{-1}(y))$$

Figure 4. Comparison of $\tilde{s}$ and $\hat{s}$



Notes: The figure plots firms' expectations of worker ability without LLMs, $\hat{s}(y)$, and when LLMs are available, $\tilde{s}(y)$. Since LLMs improve signal quality, firms know that low signals must come from non-users, so the two expectations coincide for small y.

The total value in this economy with imperfect information and LLMs is given by,

$$V_L = \int_{\mathcal{S}} \int_{\mathcal{S}} f(G_x^{-1}(G_{\tilde{s}}(\tilde{s})), s) \mathrm{d}G_{\tilde{s}|s}(\tilde{s}|s) \mathrm{d}G_s(s) \tag{9}$$

Proposition 1. *The total value in the economy with imperfect information and LLMs is lower than the total value of an economy with only imperfect information.*

Proposition 1 states that the aggregate value in the economy with imperfect information and LLMs (V_L) is at most equal to the value in the economy with imperfect information without LLMs.[30] This result reflects the fact that while LLMs can enhance the signals for some workers, the ambiguity created by the inability of firms to distinguish augmented signals from non-augmented ones leads to systematically lower worker ability estimates for certain signal ranges. This downward bias in estimates causes suboptimal matches, particularly due to the concavity and complementarity of the match value function f, thereby reducing

[30]See Appendix E for details and proof.

overall output. Hence, since LLMs reduce individual disparities in signals, their aggregate effect will still fail to surpass the output of the imperfect information economy without LLM assistance.

5.1 Calibration

To quantify the efficiency losses as a result of LLMs with imperfect information, we calibrate parameters of the model and simulate the output losses. We assume a CES function for the value generated by the match of a firm and a worker, following Eeckhout and Kircher (2018), $f(x, s) = [\alpha x^\rho + (1-\alpha)s^\rho]^{1/\rho}$. Table 8 reports the parameters and the moments used in the calibration.

Figure 5 plots the efficiency loss as a result of LLMs on the vertical axis, as a percentage with respect to the efficiency in the case of positively assortative matching (PAM) with imperfect information but *without* LLMs (denoted by V_I). On the horizontal axis, p (the proportion of job-seekers that do *not* use LLMs) ranges from 0 to 1. Efficiency losses can amount to 0.8%, when approximately 80% of job-seekers use LLMs. In the job-seeker and recruiter experiments, the proportion of cover letters written with LLM assistance were 47% and 50%, respectively, which would result in approximately 0.4% efficiency losses.

One important feature of Figure 5 is that after a certain threshold of proportion of users of LLMs (in our calibration 80%), efficiency losses, while still positive, begin to decline. This is because as a very high percentage of applicants use LLMs, the uncertainty caused by the firm's lack of information about LLMs usage begins to decrease, thus reducing the probability of mismatches. In the extreme when all applicants use LLMs ($p = 1$), this uncertainty is fully removed and there are no efficiency losses. This is because in this scenario firms know with certainty that every cover letter has been written with LLMs, and thus invert the $h(\cdot)$ function for all signals to obtain the same estimate of worker quality as in the case where nobody was using LLMs, i.e. $\tilde{s}(y) = \hat{s}(y)$ for all y.

Figure 5 also shows that the LLM-induced misallocation is not symmetric around 0.5 with respect to the proportion of job-seekers that use LLMs. This asymmetry is due to the distortionary effect of LLMs on the cover letter signals. By shifting the left tail of signals upwards, and having little effect on the right tail (as motivated by our second experiment), it effectively causes the distribution of signals to become more concentrated, thus less informative to firms as a signal of quality. This distortion only increases as the proportion of applicants who use LLMs increases (lower p). The negative effect on output only begins to decline when the proportion of LLM users is so high that they are effectively sure that most applicants are using LLMs and so they take this into account in their evaluations.

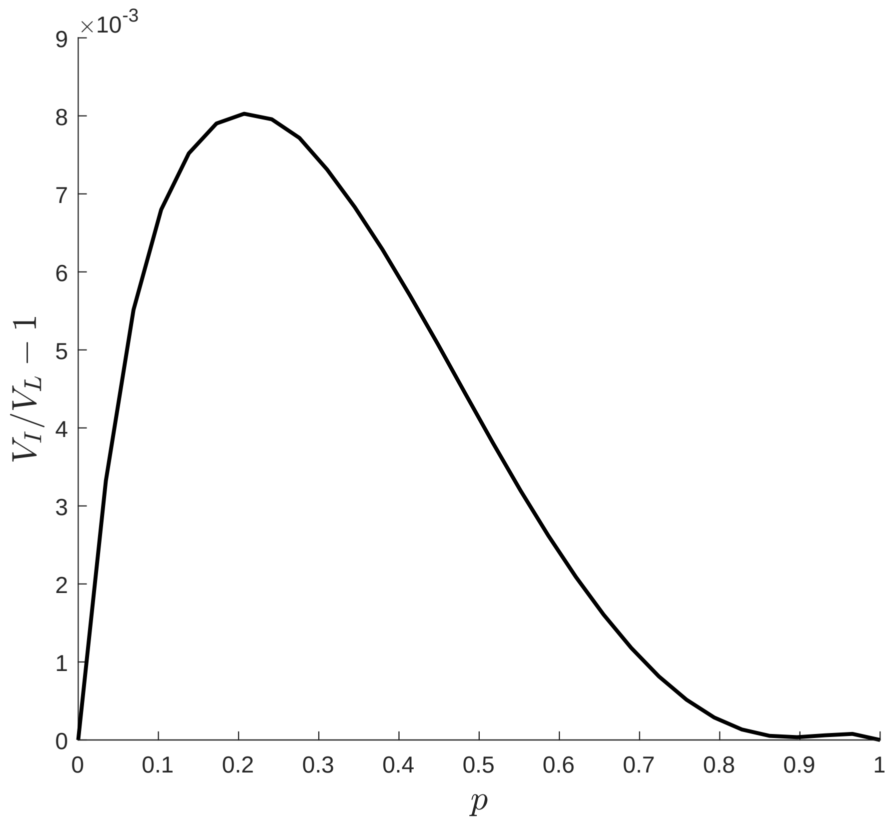
In an extension of the model presented in Appendix E we present a model where the use of LLM is positively or negatively correlated with underlying ability s. This could be, for instance, due to higher ability applicants being more open to using, or having better access to, new technologies such as AI; or the opposite, due to lower qualified applicants anticipating the higher effectiveness of LLMs on their cover letters and thus choosing to use it more intensely. We do this by making p a linear function of underlying quality s and varying its slope, while adjusting its intercept to keep the *average* probability of adoption across all applicants constant at 0.5. Our results show that total output increases with an increasing correlation of LLM adoption with underlying ability. This is in line with our previous results about the model: a higher, more positive, correlation with ability means that more applicants with higher s use the LLM and less applicants with lower s use it. Since the LLM has a larger effect on applicants with lower s this means that fewer signals in the cross-section of applicants are distorted, thus leading to fewer mismatches on the labor market.

Table 8: Calibration Parameters

Parameter	Source	Value
Production Function	Eeckhout and Kircher (2018)	$\alpha = 0.5; \rho = 0.5$
Distribution of Worker Type	Teulings (1995)	$N(5.59, 1.29)$
Distribution of Firms Type	Teulings (1995)	truncated $N(5.59, 1.29)$
Slope of h function	Experiment #1	0.61

Notes: From the control group's cover letters in Experiment #1, we have $\mathbb{E}[y_1] = 5.59$ and $Var[y_1] = 2.61$. Since $y_1 = s + e$ and e is assumed to have 0 mean, we get $E[s] = 5.59$, which we use to calibrate the moments of the workers' type distribution. In this calibration we set the firm's type distribution the same as the workers' but trancated to the largest non-negative interval around the mean, i.e. $[0, 11.18]$ in order to avoid negative values. To get the slope of the h function we first calculate the quantile treatment effect (Figure 1) for each 5-th percentile, then we regress the percentiles on the treatment effects and a constant. The slope of the h function is then one minus the estimated slope coefficient, which is -0.39.

Figure 5. Sensitivity of Misallocation to p



Notes: The figure shows how the total output in the economy with the presence of LLM technology V_L (calculated using equation (9)) differs from the economy where there are imperfect labor market signals but no LLM technology V_I (see Appendix E for details) when the proportion p of applicants who do *not* use LLM technology varies over $[0, 1]$.

6 Conclusion

The emergence of Large Language Models has the potential to significantly impact productivity and reshaped labor market dynamics. LLMs are widely adopted, with Bick et al. (2025) documenting that nearly 40% of the US working-age population is using them, and 23% of employees using them at least once at work in the last week. However, the influence of their use by job-seekers on labor market signals and subsequent matching efficiency remains largely unexplored. This paper addresses this gap by conducting two field experiments involving job-seekers and recruiters, the findings of which inform a calibration exercise of an assignment model with incomplete information and LLM usage.

In our job-seekers' experiment, we find that access to OpenAI's ChatGPT improves the average quality of cover letters. However, these gains do not translate to improved chances of receiving a job interview, as the ChatGPT-induced improvements are limited to less critical and personalized sections of the cover letter. Furthermore, while access to ChatGPT reduces the frequency of spelling and grammar errors, it also reduces the cover letter's readability by using longer words. To understand how recruiters perceive job-seekers' use of this technology, our second experiment asked recruiters to evaluate cover letters while varying the degree of disclosure regarding whether ChatGPT was used. On average, we find no significant difference in evaluations based on the degree of disclosure. However, high-quality cover letters written without ChatGPT were evaluated more positively when recruiters were aware of this, suggesting that recruiters place a premium on high-quality applications when it is evident that the cover letter was written without LLM assistance.

Combined, our experimental results reveal important non-linearities in how LLMs impact job-seekers. Lower-ability candidates clearly benefit from LLM assistance, whereas applicants already producing high-quality cover letters experience little improvement or may even be disadvantaged if recruiters become cautious about AI-generated content. These nuanced effects inform our model, which examines the broader welfare implications of widespread LLM adoption under conditions consistent with our experimental observations. Within this model and calibration, we find that LLM usage consistently reduces overall welfare by distorting labor market matching: lower-ability applicants appear more skilled than they are and get placed in roles demanding higher qualifications, while high-quality applicants are mismatched into positions for which they are overqualified. Although some firms seeking lower-skilled employees might occasionally benefit from this misalignment, the net effect is a reduction in aggregate welfare relative to a baseline scenario with imperfect information but no LLM assistance.

The findings from the two experiments and calibration are important, given that generative AI is still in its early stages, with limited regulation and uncertain future norms. For example, if regulations mandating explicit disclosure of LLM use—such as those currently being studied by the European Union—become widespread, the conditions in our *Full In-*

formation treatment may resemble future reality.[31] Insights from this paper are likely to extend beyond the labor market, to situations in which a decision-maker cannot fully observe an applicant's ability. One such example is university admission essays. Universities have taken different approaches to the use of LLMs in admissions applications, ranging from bans to guidance on how to use them.[32] While the domain is different from our setting, the underlying characteristics—applicants submitting written texts outlining their motivation and qualification, and evaluators not knowing whether LLMs were used or not—are very similar. As such, our findings, in particular the heterogeneous effects of LLM assistance on the signal's quality and the importance of recruiters' knowledge and perception of LLM usage, can extend to other contexts. This is particularly important during the current transitional phase with LLMs, where models are constantly improving, and individuals and firms are still understanding use cases, best practices, and policies in response to them.

Given the accelerating capabilities of generative AI—our study employs the freely accessible ChatGPT 3.5, which is already outdated by newer, more advanced models—our results likely represent a conservative estimate of the technology's future impact. Therefore, the impact of generative AI on labor market signals and matching presents numerous promising avenues for future research, as our results suggest that the winners and losers can depend on the user's innate ability, perceptions of evaluators, and disclosure policies.

[31] Several academic journals and university admissions teams already require AI and LLM disclosure statements.

[32] The University of Michigan Law School bans AI tools in their applications, while Arizona State University Law School allows applicants to use them as long as they disclose them, and Georgia Tech offers AI guidance to applicants (The Guardian, 2023).

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2022). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1):1–35.

Acemoglu, D. (2024). The simple macroeconomics of AI. Technical report, MIT Department of Economics.

Almgren, F. J. and Lieb, E. H. (1989). Symmetric decreasing rearrangement is sometimes continuous. *Journal of the American Mathematical Society*, 2(4):683–773.

Alsudais, A. (2025). Exploring the change in scientific readability following the release of ChatGPT. *Journal of Informetrics*, 19(3):101679.

Ameriks, J., Briggs, J., Caplin, A., Shapiro, M. D., and Tonetti, C. (2020). Long-term-care utility and late-in-life saving. *Journal of Political Economy*, 128(6):2375–2451.

Andre, P., Pizzinelli, C., Roth, C., and Wohlfart, J. (2022). Subjective models of the macroeconomy: Evidence from experts and representative samples. *The Review of Economic Studies*, 89(6):2958–2991.

ANP (2023). Kwart organisaties screent sociale media van sollicitant. Accessed: 2024-11-25.

ANP (2024). Werkgevers gedogen sollicitatiebrief en cv van ChatGPT. Accessed: 2024-11-25.

Arcidiacono, P., Bayer, P., and Hizmo, A. (2010). Beyond Signaling and Human Capital: Education and the Revelation of Ability. *American Economic Journal: Applied Economics*, 2(4):76–104.

Arellano-Bover, J. (2024). Career consequences of firm heterogeneity for young workers: First job and firm size. *Journal of Labor Economics*, 42(2):549–589.

Autor, D. (2024). Applying AI to rebuild middle-class jobs. Technical Report 32140, National Bureau of Economic Research.

Autor, D. H. (2001). Wiring the labor market. *Journal of Economic Perspectives*, 15(1):25–40.

Avery, M., Leibbrandt, A., and Vecci, J. (2023). Does artificial intelligence help or hurt gender diversity? Evidence from two field experiments on recruitment in tech. Technical report, Monash University. Available at SSRN: https://ssrn.com/abstract=4370805.

Avery, M., Leibbrandt, A., and Vecci, J. (2024). Does artificial intelligence help or hurt gender diversity? Evidence from two field experiments on recruitment in tech. CESifo Working Paper.

Awuah, K., Krenk, U., and Yanagizawa-Drott, D. (2025). Automation with generative AI? Evidence from a teacher hiring pipeline. Working Paper.

Bashardoust, A., Feng, Y., Geissler, D., Feuerriegel, S., and Shrestha, Y. R. (2024). The effect of education in prompt engineering: Evidence from journalists.

Bassi, V. and Nansamba, A. (2021). Screening and signalling non-cognitive skills: Experimental evidence from Uganda. *The Economic Journal*, 132(642):471–511.

Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political economy*, 81(4):813–846.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29–50.

Bick, A., Blandin, A., and Deming, D. J. (2025). The rapid adoption of generative AI. Working Paper.

Böhm, R., Jörling, M., Reiter, L., et al. (2023). People devalue generative AI's competence but not its advice in addressing societal and personal challenges. *Communications Psychology*, 1:32.

Bohren, N., Hakimov, R., and Lalive, R. (2024). Creative and strategic capabilities of generative AI: Evidence from large-scale experiments. Technical Report DP No. 17302, IZA Institute of Labor Economics, Bonn, Germany. IZA Discussion Paper Series; ISSN: 2365-9793.

Brañas-Garza, P., Estepa-Mohedano, L., Jorrat, D., Orozco, V., and Rascón-Ramírez, E. (2021). To pay or not to pay: Measuring risk preferences in lab and field. *Judgment and Decision Making*, 16(5):1290–1313.

Brañas-Garza, P., Jorrat, D., Espín, A. M., and Sánchez, A. (2023). Paid and hypothetical time preferences are the same: lab, field and online evidence. *Experimental Economics*, 26(2):412–434.

Brock, F. (2000). A general rearrangement inequality à la Hardy–Littlewood. *Journal of Inequalities and Applications*, 2000(4):805282.

Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 1(4):200–232.

Brynjolfsson, E., Li, D., and Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics*.

Burchard, A. (2009). A short course on rearrangement inequalities. *Lecture notes, IMDEA Winter School, Madrid*.

Burchard, A. and Hajaiej, H. (2006). Rearrangement inequalities for functionals with monotone integrands. *Journal of Functional Analysis*, 233(2):561–582.

Caplin, A., Deming, D. J., Li, S., Martin, D. J., Marx, P., Weidmann, B., and Ye, K. J. (2024). The ABC's of who benefits from working with AI: Ability, beliefs, and calibration. Technical report, National Bureau of Economic Research.

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J.-F., Brañas-Garza, P., Butera, L., Douglas, K. M., Everett, J. A. C., Gigerenzer, G., Greenhow, C., Hashimoto, D. A., Holt-Lunstad, J., Jetten, J., Johnson, S., Kunz, W. H., Longoni, C., Lunn, P., Natale, S., Paluch, S., Rahwan, I., Selwyn, N., Singh, V., Suri, S., Sutcliffe, J., Tomlinson, J., van der Linden, S., Van Lange, P. A. M., Wall, F., Van Bavel, J. J., and Viale, R. (2024). The impact of generative artificial intelligence on socioeconomic inequalities and policy making. *PNAS Nexus*, 3(6):pgae191.

Carranza, E., Garlick, R., Orkin, K., and Rankin, N. (2022). Job search and hiring with limited information about workseekers' skills. *American Economic Review*, 112(11):3547–83.

Carvajal, D., Franco, C., and Isaksson, S. (2024). Will artificial intelligence get in the way of achieving gender equality? Discussion Paper 03, NHH Dept. of Economics. The paper has been revised in the Paper Series, and can be found here: https://openaccess.nhh.no/nhh-xmlui/handle/11250/3122396.

Chan, W., Hao, L., and Suen, W. (2007). A sitnaling theory of grade inflation. *International Economic Review*, 48(3):1065–1090.

Chaturvedi, S. and Chaturvedi, R. (2025). Who gets the callback? Generative AI and gender bias. *arXiv preprint arXiv:2504.21400*.

Choi, J. H. and Schwarcz, D. (2024). AI assistance in legal analysis: An empirical study. *Journal of Legal Education*. Available at SSRN: https://ssrn.com/abstract=4539836 or http://dx.doi.org/10.2139/ssrn.4539836.

Criddle, C. and Strauss, D. (2024). Jobhunters flood recruiters with AI-generated CVs. https://www.ft.com/content/30a032dd-bdaa-4aee-bc51-754867abbde0. Financial Times, accessed on [14/09/2024].

Crowe, J., Zweibel, J., and Rosenbloom, P. (1986). Rearrangements of functions. *Journal of functional analysis*, 66(3):432–438.

Cui, Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S., and Salz, T. (2025). The effects of generative AI on high-skilled work: Evidence from three field experiments with software developers. *SSRN Electronic Journal*. Available at SSRN.

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Available at SSRN: https://ssrn.com/abstract=4573321 or http://dx.doi.org/10.2139/ssrn.4573321.

Dell'Acqua, F. (2022). Falling asleep at the wheel: Human/AI collaboration in a field experiment on hr recruiters. Working paper. Unpublished manuscript.

Deming, D. and Kahn, L. B. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, 36(S1):S337–S369.

Deming, D. J. (2017). The growing importance of social skills in the labor market*. *The Quarterly Journal of Economics*, 132(4):1593–1640.

Deming, D. J., Ong, C., and Summers, L. H. (2025). Technological disruption in the labor market. Working Paper 33323, National Bureau of Economic Research.

Eeckhout, J. and Kircher, P. (2018). Assortative matching with large firms. *Econometrica*, 86(1):85–132.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). Gpts are gpts: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.

Englmaier, F., Grimm, S., Grothe, D., Schindler, D., and Schudy, S. (2024). The effect of incentives in nonroutine analytical team tasks. *Journal of Political Economy*, 132(8):2695–2747.

Eric Reicin (2021). AI can be a force for good in recruiting and hiring new employees. Accessed: 2024-12-08.

Falk, A., Becker, A., Dohmen, T., Huffman, D., and Sunde, U. (2023). The preference survey module: A validated instrument for measuring risk,time, and social preferences. *Management Science*, 69(4):1935–1950.

Filippucci, F., Gal, P., and Schief, M. (2024). Miracle or myth? Assessing the macroeconomic productivity gains from artificial intelligence. Technical Report 29, OECD Publishing.

Hoffman, M., Kahn, L. B., and Li, D. (2017). Discretion in hiring. *The Quarterly Journal of Economics*, 133(2):765–800.

Institute for the Future of Work (2022). Algorithmic hiring systems. Accessed: 2024-12-08.

Kadoma, K., Metaxa, D., and Naaman, M. (2024). Generative AI and perceptual harms: Who's suspected of using LLMs?

Kurlat, P. and Scheuer, F. (2020). Signalling to experts. *The Review of Economic Studies*, 88(2):800–850.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102.

Merali, A. (2024). Scaling laws for economic productivity: Experimental evidence in LLM-assisted translation.

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

Otis, N., Clarke, R., Delecourt, S., Holtz, D., and Koning, R. (2024). The uneven impact of generative AI on entrepreneurial performance. *Management Science*. Available at SSRN: https://ssrn.com/abstract=4671369 or http://dx.doi.org/10.2139/ssrn.4671369.

Peng, B., Galley, M., He, P., Cheng, H., Xie, Y., Hu, Y., Huang, Q., Liden, L., Yu, Z., Chen, W., and Gao, J. (2023a). Check your facts and try again: Improving large language models with external knowledge and automated feedback.

Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023b). The impact of AI on developer productivity: Evidence from GitHub Copilot.

Rendement (2024). AI-chatbots veranderen het speelveld bij sollicitaties. https://www.rendement.nl/werving-en-selectie/verdiepingsartikel/ai-chatbots-veranderen-het-speelveld-bij-sollicitaties.html. Accessed: 2025-07-11.

Resume Genius (2024). 50+ cover letter statistics for 2025 (hiring manager survey). https://resumegenius.com/blog/cover-letter-help/cover-letter-statistics. Accessed: 2025-07-11.

Schwager, R. (2012). Grade inflation, social background, and labour market matching. *Journal of Economic Behavior & Organization*, 82(1):56–66.

Schwartzberg, J. (2025). Cover letters still matter—even if they're not required. *Harvard Business Review*. Accessed: 2025-07-11.

Shanahan, M. (2024). Talking about large language models. *Communications of the ACM*, 67(2):68–79.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374.

Stango, V. and Zinman, J. (2023). We are all behavioural, more, or less: A taxonomy of consumer decision-making. *The Review of Economic Studies*, 90(3):1470–1498.

Teulings, C. N. (1995). The wage distribution in a model of the assignment of skills to jobs. *Journal of Political Economy*, 103(2):280–315.

The Guardian (2023). ChatGPT and AI: How they can help disadvantaged college applicants in the era of affirmative action.

The White House (2022). The impact of artificial intelligence on the future of workforces in the european union and the united states of america. Technical report, The White House.

United States. Bureau of the Census and United States. Bureau of Labor Statistics (2021). Current population survey, May 2017: Contingent worker supplement.

U.S. Bureau of Labor Statistics (2023). Work statistics: News release. Accessed: 2024-12-08.

Wiles, E., Munyikwa, Z., and Horton, J. (2025). Algorithmic writing assistance on jobseekers' resumes increases hires. *Management Science*.

Zety (2024). Zety report finds 81% of recruiters have rejected a candidate based on details in their cover letter. https://zety.com/blog/recruiting-preferences. Accessed: 2025-07-11.

Zhang, Y. and Gosline, R. (2023). Human favoritism, not AI aversion: People's perceptions (and bias) toward generative AI, human experts, and human–GAI collaboration in persuasive content generation. *Judgment and Decision Making*, 18:e41.

Online Appendix to:

Labor Market Signals: The Role of Large Language Models

by Kian Abbas Nejad, Giuseppe Musillo, Till Wicker, and Niccolò Zaccaria

A Figures and Tables

A.1 Job-seeker Experiment

A.1.1 Balance Table

Table A1: Balance Table: Demographics

Variable	N	(1) ***Control*** Mean/(SD)	N	(2) ***Treatment*** Mean/(SD)	(1)-(2) Pairwise t-test P-value
Dutch Speaker	72	0.236	65	0.200	0.613
		(0.428)		(0.403)	
GPA	72	7.505	65	7.557	0.644
		(0.669)		(0.639)	
English Ability	72	6.069	65	6.138	0.597
		(0.738)		(0.788)	
Master's Student	72	0.528	65	0.585	0.507
		(0.503)		(0.497)	
Age	72	23.597	65	23.123	0.457
		(4.023)		(3.347)	
Econ or Business Degree	72	0.750	65	0.738	0.878
		(0.436)		(0.443)	
Tilburg University	72	0.722	65	0.677	0.567
		(0.451)		(0.471)	
Female	72	0.486	65	0.538	0.544
		(0.503)		(0.502)	

Notes: Dutch speaker is a dummy variable equal to one if the student self-reports that they speak Dutch. GPA is on a scale from 0-10, and self-reported. English ability is on a scale from 1-7. Master's student is a dummy variable if the student is enrolled in a Master's program. Age is the student's age. Econ or Business Degree is a dummy variable if the student is enrolled at Tilburg's School of Economics and Management, or Utrecht's School of Economics. Tilburg University is a dummy equal to one if the student is enrolled at Tilburg University. Female is a dummy equal to one if the student identifies as a female.

Table A2: Balance Table: CV Evaluation

	(1)		(2)		(1)-(2)
	Control		***Treatment***		Pairwise t-test
Variable	N	Mean/(SD)	N	Mean/(SD)	P-value
CV Grade	144	5.777	130	5.428	0.075*
		(1.567)		(1.663)	
CV: Layout	144	0.017	130	-0.019	0.768
		(0.997)		(1.007)	
CV: Education	144	0.107	130	-0.119	0.062*
		(1.000)		(0.991)	
CV: Experience	144	0.102	130	-0.113	0.077*
		(0.960)		(1.034)	
CV: Extra Curricular	144	0.081	130	-0.089	0.161
		(0.944)		(1.055)	

Notes: The sub-components of the CV are described in more detail in Appendix B.1.3. Evaluations of the sections are on a scale from 0-10.

A.1.2 Heterogeneity - By Gender

Table A3: Heterogeneous Results: Men vs. Women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter			Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
ChatGPT	0.155	0.256	0.196	0.059	-0.030	0.211	0.194	0.047
	(0.176)	(0.208)	(0.191)	(0.153)	(0.185)	(0.193)	(0.354)	(0.522)
Female	-0.022	0.109	0.001	-0.026	-0.205	0.078	0.127	0.207
	(0.170)	(0.180)	(0.177)	(0.170)	(0.159)	(0.197)	(0.414)	(0.617)
ChatGPT*Female	0.123	-0.215	0.072	0.058	0.335	0.170	0.035	-0.087
	(0.236)	(0.276)	(0.239)	(0.222)	(0.237)	(0.244)	(0.577)	(0.875)
Control Group Mean	0.000	0.000	0.000	0.000	0.000	0.000	2.799	0.278
Observations	274	274	274	274	274	274	274	274

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Standard errors are clustered at the individual level. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Column 1-7 refer to variables as described in Appendix B.1.3, while Colummn 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Female is a dummy variable equal to one if the job-seeker identifies as a female, and zero otherwise. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A.1.3 Heterogeneity - By Economics Degree vs. Not

Table A4: Heterogeneous Results: Econ vs. Non-Econ degree

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter			Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
ChatGPT	0.267	0.360	0.152	-0.085	0.238	0.408	0.083	-0.298
	(0.247)	(0.259)	(0.197)	(0.296)	(0.267)	(0.256)	(0.669)	(0.886)
Econ/Business Degree	0.078	-0.067	-0.033	-0.077	0.229	0.231	0.904*	0.861
	(0.221)	(0.220)	(0.191)	(0.267)	(0.220)	(0.230)	(0.508)	(0.785)
ChatGPT*Econ/Business	-0.066	-0.279	0.107	0.228	-0.130	-0.150	0.128	0.451
	(0.282)	(0.303)	(0.252)	(0.323)	(0.292)	(0.287)	(0.724)	(1.009)
Control Group Mean	0.000	0.000	0.000	0.000	0.000	0.000	2.799	0.278
Observations	274	274	274	274	274	274	274	274

Standard errors in parentheses

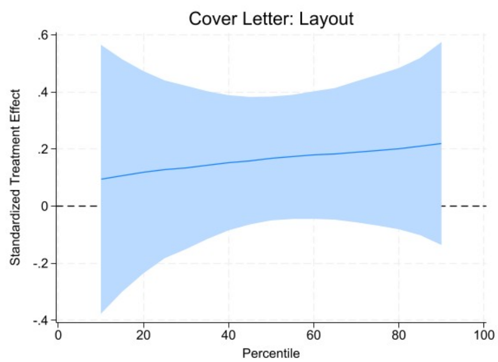
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Standard errors are clustered at the individual level. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Column 1-7 refer to variables as described in Appendix B.1.3, while Colummn 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Econ/Business is a dummy variable equal to 1 if students are enrolled in Tilburg University's School of Economics and Management, or the Utrecht School of Economics. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

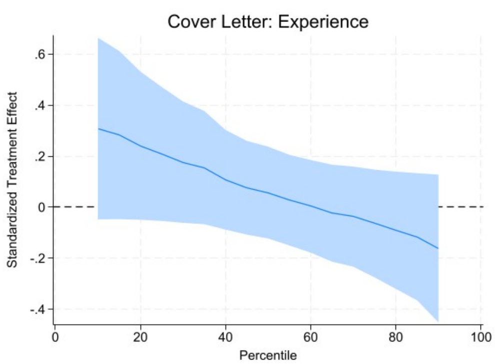
A.1.4 Heterogeneity by Ability: Layout, Motivation, Experience

Figure A1. Quantile Regressions

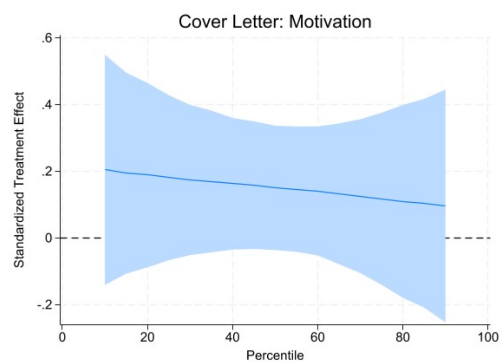
(a) Cover Letter: Layout



(b) Cover Letter: Motivation



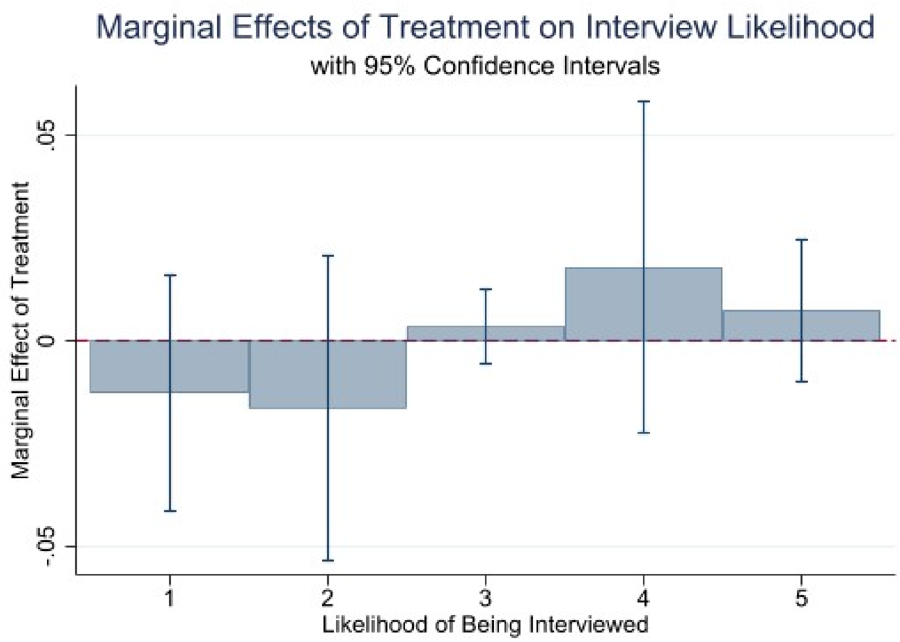
(c) Cover Letter: Experience



Notes: Each panel plots the estimated treatment effect of ChatGPT on cover letters across quantiles of the outcome distribution, obtained from quantile regressions. The outcomes correspond to the three main sub-components of the cover letter—Layout, Motivation, and Experience—as described in Appendix B.1.3. Shaded areas represent 95% confidence intervals based on robust standard errors clustered at the individual level. The corresponding average (OLS) treatment effects are reported in Table 1.

A.1.5 Marginal Treatment Effects: Likelihood of Interview

Figure A2. Marginal Treatment Effects: Likelihood of Interview



Notes: The figure shows the marginal treatment effect of being assigned to the treatment arm in Experiment #1, on the likelihood of receiving a score from 1 to 5 on the likelihood of being interviewed (higher scores are better).

A.1.6 ChatGPT as a Recruiter

Table A5: ChatGPT as the Recruiter

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ChatGPT as the Recruiter							
	Cover Letter						Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
ChatGPT	0.286***	0.219***	0.264**	0.213*	0.286**	0.377***	0.981*	2.218
	(0.099)	(0.074)	(0.133)	(0.114)	(0.116)	(0.119)	(0.530)	(2.565)
Control Group Mean	0.000	0.000	0.000	0.000	-0.000	0.000	4.028	0.875
Observations	137	137	137	137	137	137	137	128

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Standard errors are clustered at the individual level. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Column 1-7 refer to variables as described in Appendix B.1.3, while Colummn 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Column 8 contains 9 fewer observations because one of hte 'school fixed effect' dummies perfectly predicts the outcome variable, and hence those observations are dropped. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A.1.7 Robustness: Double Lasso (Belloni et al., 2014)

Table A6: Double Lasso regression

	(1)	(2)	(3)	(4)	(5)	(6)
				Cover Letter		
	Total	Layout	Intro	Experience	Motivation	Closing
ChatGPT	0.218**	0.170	0.223**	0.094	0.129	0.296**
	(0.111)	(0.138)	(0.114)	(0.105)	(0.112)	(0.117)
Control Group Mean	0.000	0.000	0.000	0.000	0.000	0.000
Observations	274	274	274	274	274	274

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions. All regressions use the PD Lasso technique to identify controls, along with recruiter and school-level fixed effects. Standard errors are clustered at the individual level. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Column 1-6 refer to variables as described in Appendix B.1.3. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A.1.8 Robustness: Bootstrap

Table A7: Bootstrap (200 reps)

	(1)	(2)	(3)	(4)	(5)	(6)
				Cover Letter		
	Total	Layout	Intro	Experience	Motivation	Closing
ChatGPT	0.222*	0.161	0.253**	0.075	0.150	0.281**
	(0.118)	(0.144)	(0.121)	(0.108)	(0.120)	(0.115)
Control Group Mean	0.000	0.000	0.000	0.000	0.000	0.000
Observations	274	274	274	274	274	274

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Standard errors are clustered at the individual level, and bootstrapped with 200 repetitions. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Column 1-6 refer to variables as described in Appendix B.1.3. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A.1.9 Time Taken

Another dimension along which job-seekers using ChatGPT and not using ChatGPT could differ is the time taken to write the cover letter. If ChatGPT substitutes ones own writing, job-seekers can write cover letters more quickly. On the contrary, if job-seekers use ChatGPT to complement their own writing, the whole process can take longer. Table A8 reports average treatment effects for the time taken to write cover letters, with the use of ChatGPT reducing the amount of time job-seekers spent writing a cover letter by 2 minutes on average, a result which is not statistically significant. This null result is robust to winsorizing individuals that wrote for more than 60 and 75 minutes, the recommended time job-seekers had.

Table A8: Time Taken to Write Cover Letter

	(1)	(2)	(3)
		Time Taken	
	No wins.	Wins. 75 min	Wins. 60 min
ChatGPT	-2.120	-1.568	-0.327
	(1.815)	(1.679)	(1.422)
Control Group Mean	57.13	56.62	54.88
Observations	132	132	132

Notes: Intention to Treat estimates from OLS regressions. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Standard errors are clustered at the individual level. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Time taken is the amount of time students took to write the cover letter. Columns 2 and 3 winsorize the time taken at 75 and 60 minutes, respectively. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A.1.10 Browser History

Balance Table: Those with vs. without browser history

Table A9: Balance Table: Browser History

	(1)		(2)		(1)-(2)
	Has Browser History		***No Browser History***		Pairwise t-test
Variable	N	Mean/(SD)	N	Mean/(SD)	P-value
Cover Letter Grade	119	-0.049 (1.070)	18	0.083 (0.960)	0.621
CV Grade	119	5.636 (1.670)	18	5.500 (1.727)	0.749
Assignment to Treatment	119	0.521 (0.502)	18	0.167 (0.383)	0.005***
Speaks Dutch	119	0.202 (0.403)	18	0.333 (0.485)	0.211
GPA	119	7.570 (0.640)	18	7.263 (0.695)	0.063*
English Proficiency	119	6.118 (0.750)	18	6.000 (0.840)	0.542
Master's Student	119	1.546 (0.500)	18	1.611 (0.502)	0.609
Age	119	23.101 (3.583)	18	25.167 (4.148)	0.027**
Econ/Business Degree	119	0.748 (0.436)	18	0.722 (0.461)	0.818
At Tilburg University	119	0.689 (0.465)	18	0.778 (0.428)	0.447
Female	119	0.521 (0.502)	18	0.444 (0.511)	0.548

Notes: Cover Letter and CV Grade are the average cover letter and CV grades, as evaluated by the recruiters. Assignment to Treatment is a dummy equal to one if the student was assigned to the ChatGPT treatment. Dutch speaker is a dummy variable equal to one if the student self-reports that they speak Dutch. GPA is on a scale from 0-10, and self-reported. English ability is on a scale from 1-7. Master's student is a dummy variable if the student is enrolled in a Master's program. Age is the student's age. Econ or Business Degree is a dummy variable if the student is enrolled at Tilburg's School of Economics and Management, or Utrecht's School of Economics. Tilburg University is a dummy equal to one if the student is enrolled at Tilburg University. Female is a dummy equal to one if the student identifies as a female.

Lee Bounds (2009) Calculation

There is differential 'attrition' of the browser history across treatments. In the Control group, we have the browser history for 79.17% of the sample. In the Treatment group, we have the browser history for 95.38% of the sample. The differential attrition rate is 95.38% - 79.17% = 16.22%. This is equal to 16.22% / 95.38% = 17.00% of the Control Group sample.

To get a lower bound, we trim the top 17.00% of the control group (for each outcome variable). To get an upper bound, we trim the bottom 17.00% of the control group (for each outcome variable).

Table A10: Probability of Not Having Browser History

	(1) No Browser History
ChatGPT	-0.137**
	(0.053)
Control Group Mean	0.208
Observations	137

Notes: Intention to Treat estimates from OLS regressions. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. No Browser History is a dummy variable equal to 1 if we were unable to obtain the browser history of the student, and zero otherwise. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A11: Browser History

	(1) # Websites Visited	(2) # Google Searches	(3) Searched: Firm	(4) Searched: Cover Letter	(5) Searched: Grammar	(6) Searched: Translation
Panel A. Lee (2009) Upper Bound						
ChatGPT	-2.220	-1.939*	1.327*	0.545	-0.958**	0.486
	(1.425)	(0.906)	(0.552)	(0.363)	(0.317)	(0.268)
Control Group Mean	19.351	10.000	5.281	2.368	2.895	2.175
Observations	110	110	111	111	112	110
Panel B. Lee (2009) Lower Bound						
ChatGPT	-9.484***	-6.341***	-1.825	-1.140	-2.427***	-1.280
	(2.247)	(1.365)	(1.169)	(0.692)	(0.631)	(0.749)
Control Group Mean	19.351	10.000	5.281	2.368	2.895	2.175
Observations	112	111	119	119	119	119

Notes: Intention to Treat estimates. Column 1 reports treatment treatment effects on the total number of websites visited, while Column 2 refers to the number of google searches. Columns 3-6 refer to whether the subject searched the relevant topic. Treatment effects are reported from OLS regressions. Standard errors are clustered at the individual level. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Panels A and B correct for Lee (2009) bounds, as explained in Appendix A.1.10. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A.1.11 No Clustered Standard Errors

Table A12: Main Results: Effect of ChatGPT on Cover Letter

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter			Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
ChatGPT	0.222**	0.161	0.253**	0.075	0.150	0.281***	0.249	0.011
	(0.095)	(0.116)	(0.099)	(0.098)	(0.100)	(0.092)	(0.259)	(0.414)
Control Group Mean	0.000	0.000	0.000	0.000	0.000	0.000	2.799	0.278
Observations	274	274	274	274	274	274	274	274

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Standard errors are robust. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A.1.12 Correlation between Prompts and Cover Letter Quality

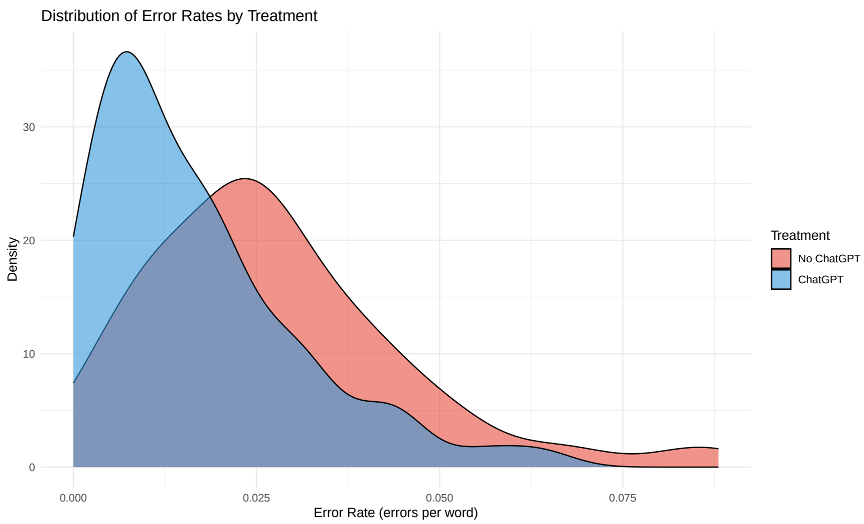
Table A13: OLS Regression: ChatGPT Prompts and Cover Letter Quality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter			Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
Number of ChatGPT Prompts	0.016	0.017	0.022	-0.023	0.016	0.032*	0.008	0.094
	(0.018)	(0.026)	(0.016)	(0.016)	(0.022)	(0.018)	(0.071)	(0.104)
Observations	94	94	94	94	94	94	94	94

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Standard errors are clustered at the individual level. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Observations refer to individuals assigned to the treatment group, for whom LLM conversation histories are available. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

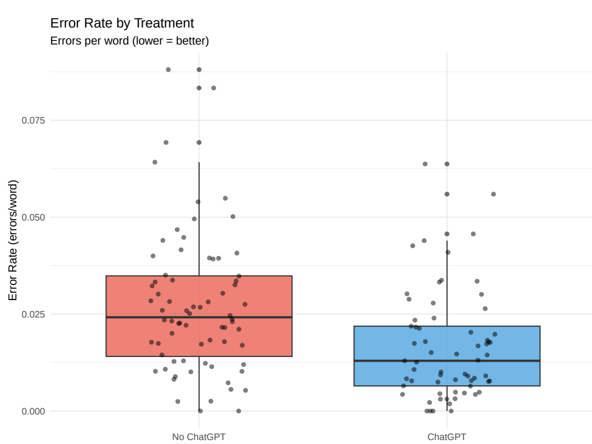
A.1.13 Readability and Error Analysis

Figure A3. Distribution of Error Rates by Treatment

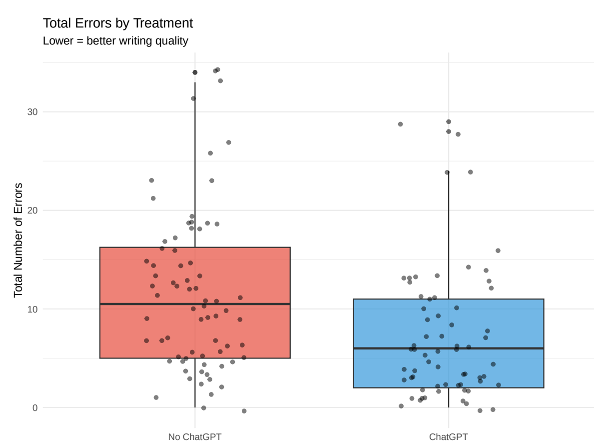


Notes: The figure shows the distribution of overall error rates (errors per word) by treatment group. ChatGPT systematically shifts the distribution toward lower error rates.

Figure A4. Technical Writing Quality by Treatment Group



(a) Error Rate by Treatment



(b) Total Errors by Treatment

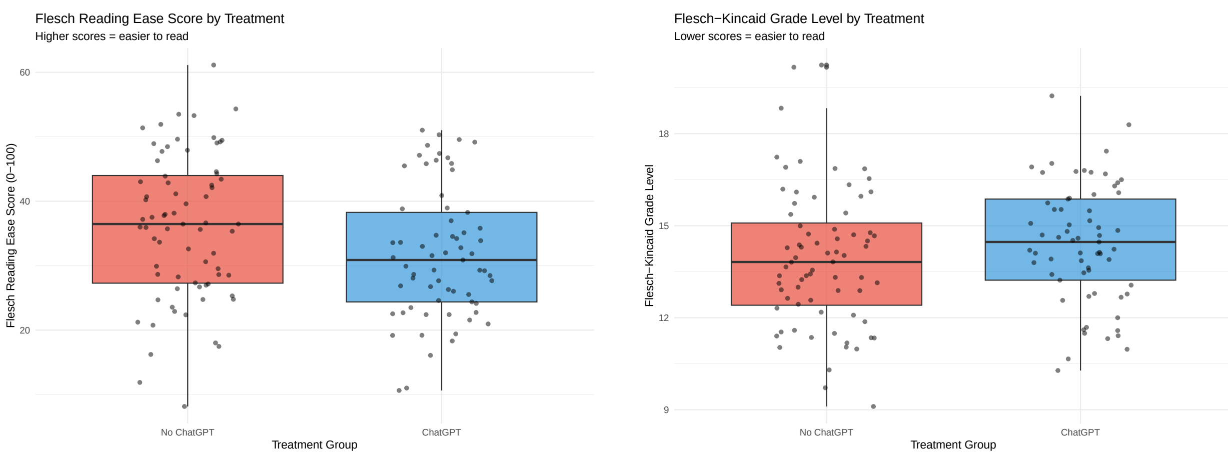
Notes: The figure shows box plots of writing error metrics by treatment group. Error rate is calculated as total errors divided by word count. ChatGPT significantly reduces both error rates and total error counts across cover letters, with medium effect sizes (Cohen's $d > 0.6$).

Table A14: Effect of ChatGPT on Technical Writing Errors

	(1) Capitalization	(2) Spelling	(3) Grammar	(4) Punctuation	(5) Typography	(6) Style
ChatGPT	0.049	-3.406***	-0.290	-0.236	0.000	-0.304**
	(0.045)	(1.044)	(0.206)	(0.193)	(.)	(0.138)
Control Group Mean	0.028	9.000	0.806	0.681	0.000	0.625
Observations	137	137	137	137	137	137
	(7) Misc.	**(8) Redundant Phrases**	**(9) Non-standard Phrases**	**(10) Commonly Confused Words**	**(11) Collocations**	**(12) Semantics**
ChatGPT	-0.221	0.000	0.000	0.000	0.000	0.050
	(0.280)	(.)	(.)	(.)	(.)	(0.053)
Control Group Mean	0.722	0.000	0.000	0.000	0.000	0.069
Observations	137	137	137	137	137	137

Notes:
Notes: Intention to Treat estimates from OLS regressions. All regressions include strata variables and imbalanced baseline variables as controls, along with recruiter and school-level fixed effects. Standard errors are clustered at the individual level. Control mean refers to the mean value of the outcome in the control group. Robust standard errors are in parentheses. Error rule categories are based on the R package 'languagetoolR ', and explained thoroughly in Wiles et al. (2025) Table A4. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

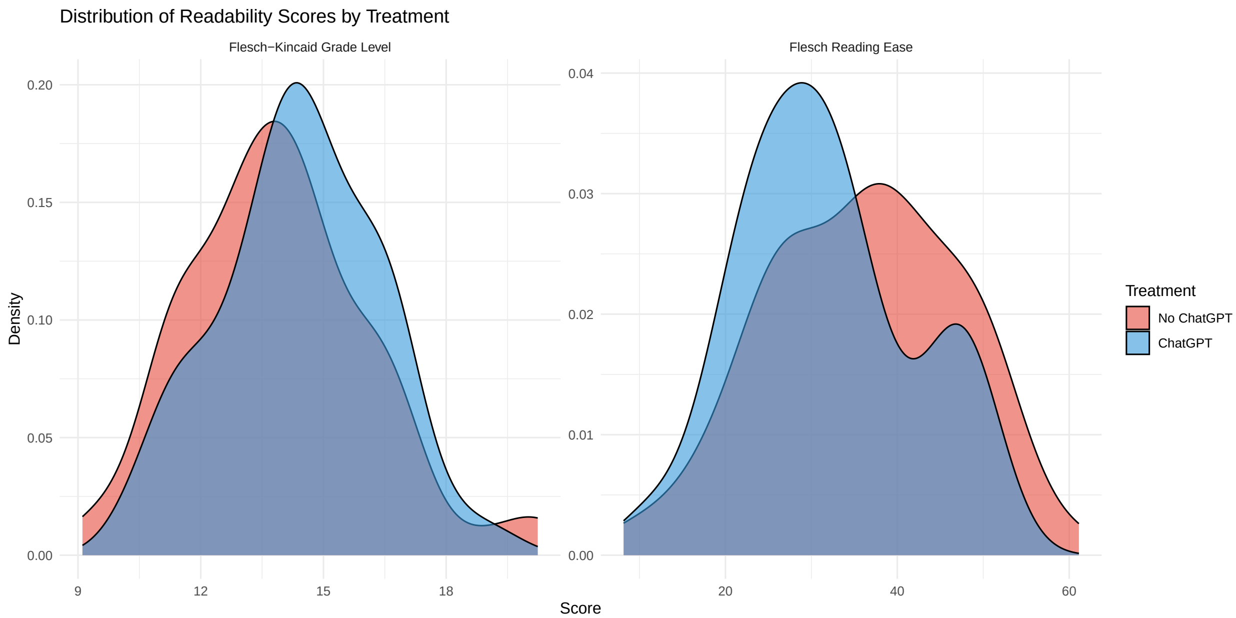
Figure A6. Readability Metrics by Treatment Group



(a) Flesch Reading Ease **(b)** Flesch-Kincaid Grade Level

Notes: The figure shows readability scores by treatment group. Higher Flesch Reading Ease scores indicate easier-to-read text, while higher Flesch-Kincaid Grade Levels indicate more complex text requiring higher education levels to understand. ChatGPT significantly reduces reading ease while increasing grade-level complexity.

Figure A5. Distribution of Readability Scores by Treatment

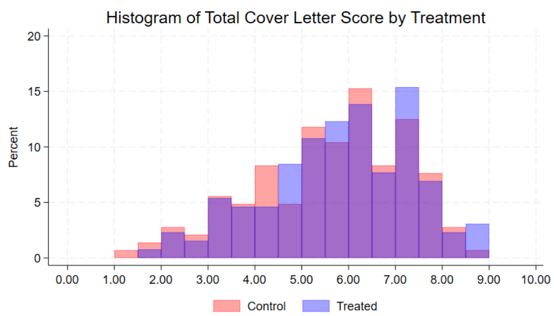


Notes: The figure shows the distribution of readability metrics by treatment group. Panel A shows Flesch Reading Ease scores (higher = easier to read). Panel B shows Flesch-Kincaid Grade Levels (higher = more complex). ChatGPT systematically shifts the distribution toward lower readability despite improving technical quality.

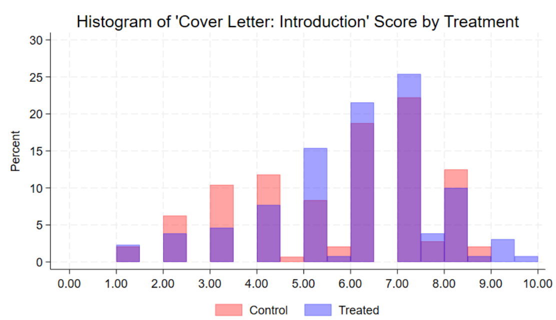
A.1.14 Histograms of Cover Letter Raw Mean Grades

Figure A7. Raw Grades, Overlapping Histograms per Treatment

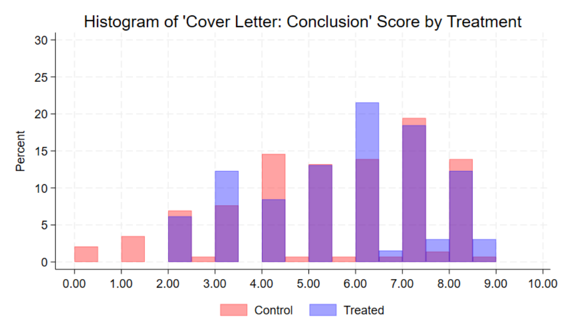
(a) Cover Letter: Total Score



(b) Cover Letter: Introduction



(c) Cover Letter: Closing



Notes: The figures show histograms of the recruiter evaluations of the job-seekers' cover letters. In particular, (a) reports the final grade; (b) reports the grade for the introduction; and (c) reports the grade for the conclusion. Blue bars reflect grades from job-seekers assigned to the treatment arm, while red bars reflect grades from job-seekers assigned to the control arm of Experiment #1.

A.2 Recruiter Experiment

A.2.1 Balance Table

Table A15: Balance Table: Demographics and Duration

		(1)		(2)		(3)	(1)-(2)	(1)-(3)	(2)-(3)
		No Info		***Partial Info***		***Full Info***	Pairwise t-test		
Variable	N	Mean/(SD)	N	Mean/(SD)	N	Mean/(SD)	P-value		
Duration (mins)	132	20.290	132	19.230	137	20.224	0.409	0.959	0.448
		(10.066)		(10.733)		(10.736)			
Intro Duration (mins)	132	3.202	132	3.120	137	3.297	0.854	0.833	0.715
		(3.312)		(3.932)		(4.013)			
Age	132	36.879	132	37.697	137	35.911	0.556	0.479	0.200
		(11.072)		(11.499)		(11.324)			
Female	132	0.598	132	0.598	137	0.562	1.000	0.547	0.547
		(0.483)		(0.492)		(0.498)			
Full-/Part-time	132	0.826	132	0.833	137	0.861	0.871	0.424	0.525
		(0.381)		(0.374)		(0.347)			
Student	132	0.250	132	0.174	137	0.248	0.133	0.973	0.139
		(0.435)		(0.381)		(0.434)			

Notes: Duration is the total number of minutes that the recruiter spent on the experiment, while Intro Duration captures the duration in minutes that the recruiter spent on the introduction. Age is the recruiter's age. Female is a dummy equal to one if the student identifies as a female. Full-/Part-time is a dummy variable equal to one if the recruiter is employed in full- or part-time employment. Student is a dummy equal to one if the recruiter is a student.

A.2.2 Additional Results

Table A16: Additional Results: Full Sample

	(1) Want to See CV	(2) Time Taken
Partial	-0.07	-0.01
	(0.17)	(0.11)
Full	-0.01	-0.01
	(0.19)	(0.10)
Cover Letter Written with GPT	0.93***	0.00
	(0.24)	(.)
Partial × Written with GPT	-0.38**	-0.05
	(0.17)	(0.11)
Full × Written with GPT	-0.12	-0.03
	(0.18)	(0.11)
t-test Partial: ChatGPT vs. No	0.75	0.27
t-test Full: ChatGPT vs. No	0.17	0.27
t-test Partial vs. Full: No ChatGPT	0.10	0.84
t-test Partial vs. Full: Yes ChatGPT	0.04	0.80
Control Group Mean	4.46	-0.00
Control Group S.D.	1.41	1.00
Observations	2005	2005

Notes: Intention to Treat estimates from OLS regressions. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses. Column (1) reports the recruiter's need to see the CV (on a 7-point Likert scale), while Column (2) reports the normalized time taken by the recruiter to read and evaluate the cover letter. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A17: Additional Results: Upper Tercile Cover Letters

	(1) Want to See CV	(2) Time Taken
Partial	-0.18	-0.17
	(0.24)	(0.14)
Full	0.11	-0.03
	(0.26)	(0.15)
Cover Letter Written with GPT	0.69***	0.00
	(0.25)	(.)
Partial × Written with GPT	-0.14	-0.02
	(0.32)	(0.17)
Full × Written with GPT	0.03	0.01
	(0.33)	(0.18)
t-test Partial: ChatGPT vs. No	0.29	0.45
t-test Full: ChatGPT vs. No	0.21	0.70
t-test Partial vs. Full: No ChatGPT	0.05	0.05
t-test Partial vs. Full: Yes ChatGPT	0.13	0.12
Control Group Mean	4.54	-0.00
Control Group S.D.	1.26	1.00
Observations	668	668

Notes: Intention to Treat estimates from OLS regressions. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses. Column (1) reports the recruiter's need to see the CV (on a 7-point Likert scale), while Column (2) reports the normalized time taken by the recruiter to read and evaluate the cover letter. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A18: Additional Results: Medium-Tercile Cover Letters

	(1) Want to See CV	(2) Time Taken
Partial	-0.33	0.11
	(0.23)	(0.16)
Full	-0.37	-0.08
	(0.26)	(0.12)
Cover Letter Written with GPT	-0.45	0.00
	(0.28)	(.)
Partial × Written with GPT	-0.03	-0.25
	(0.29)	(0.19)
Full × Written with GPT	0.34	0.03
	(0.32)	(0.16)
t-test Partial: ChatGPT vs. No	0.23	0.23
t-test Full: ChatGPT vs. No	0.58	0.40
t-test Partial vs. Full: No ChatGPT	0.41	0.62
t-test Partial vs. Full: Yes ChatGPT	0.17	0.49
Control Group Mean	4.55	-0.00
Control Group S.D.	1.33	1.00
Observations	668	668

Notes: Intention to Treat estimates from OLS regressions. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses. Column (1) reports the recruiter's need to see the CV (on a 7-point Likert scale), while Column (2) reports the normalized time taken by the recruiter to read and evaluate the cover letter. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A19: Additional Results: Low-Tercile Cover Letters

	(1) Want to See CV	(2) Time Taken
Partial	0.31	0.09
	(0.28)	(0.20)
Full	0.22	0.12
	(0.28)	(0.16)
Cover Letter Written with GPT	0.51**	0.00
	(0.26)	(.)
Partial × Written with GPT	-0.94***	0.12
	(0.30)	(0.27)
Full × Written with GPT	-0.67**	-0.18
	(0.30)	(0.19)
t-test Partial: ChatGPT vs. No	0.70	0.82
t-test Full: ChatGPT vs. No	0.52	0.02
t-test Partial vs. Full: No ChatGPT	0.85	0.40
t-test Partial vs. Full: Yes ChatGPT	0.48	0.16
Control Group Mean	4.29	0.00
Control Group S.D.	1.60	1.00
Observations	669	669

Notes: Intention to Treat estimates from OLS regressions. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses. Column (1) reports the recruiter's need to see the CV (on a 7-point Likert scale), while Column (2) reports the normalized time taken by the recruiter to read and evaluate the cover letter. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A.2.3 Recruiters Heterogeneity

Table A20: Female Recruiters: Effect of Knowledge About ChatGPT Usage on Cover Letter Evaluation

	(1) Total	(2) Layout	(3) Intro	(4) Cover Letter Experience	(5) Motivation	(6) Closing	(7) Likelihood of Interview	(8) High Chance of Interview
Partial	-0.13	-0.17	-0.13	-0.07	-0.13	-0.12	-0.20	-0.25
	(0.11)	(0.12)	(0.12)	(0.11)	(0.12)	(0.11)	(0.20)	(0.20)
Full	0.05	0.07	0.13	0.07	0.11	0.09	0.12	0.19
	(0.11)	(0.12)	(0.11)	(0.12)	(0.12)	(0.11)	(0.20)	(0.20)
Cover Letter Written with GPT	-0.05	0.04	0.02	-0.01	-0.02	-0.06	-0.14	-0.18
	(0.09)	(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	(0.18)	(0.19)
Partial × Written with GPT	0.12	0.21	0.10	0.03	0.13	0.09	0.24	0.31
	(0.13)	(0.14)	(0.14)	(0.12)	(0.13)	(0.13)	(0.24)	(0.27)
Full × Written with GPT	0.05	0.07	0.04	0.02	-0.03	0.05	-0.11	-0.21
	(0.14)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.26)	(0.27)
t-test Partial: ChatGPT vs. No	0.47	0.02	0.25	0.76	0.29	0.77	0.42	0.50
t-test Full: ChatGPT vs. No	0.96	0.30	0.57	0.90	0.59	0.93	0.26	0.06
t-test Partial vs. Full: No ChatGPT	0.03	0.03	0.00	0.04	0.02	0.01	0.25	0.21
t-test Partial vs. Full: Yes ChatGPT	0.22	0.38	0.05	0.17	0.42	0.07	0.86	0.68
Control Group Mean	0.04	0.01	0.04	0.01	0.01	0.01	3.43	0.54
Control Group S.D.	0.99	1.01	1.00	0.99	1.01	1.01	1.20	0.50
Observations	1175	1175	1175	1175	1175	1175	1175	1175

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses, and are clustered at the recruiter-level. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Sample consists of only Female recruiters. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A21: Male Recruiters: Effect of Knowledge About ChatGPT Usage on Cover Letter Evaluation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter			Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
Partial	0.14	0.03	0.20	0.09	0.11	0.05	-0.03	-0.05
	(0.14)	(0.13)	(0.12)	(0.14)	(0.15)	(0.14)	(0.24)	(0.24)
Full	0.08	0.09	0.11	0.13	0.06	0.11	0.28	0.35
	(0.14)	(0.13)	(0.13)	(0.14)	(0.14)	(0.13)	(0.24)	(0.24)
Cover Letter Written with GPT	-0.03	0.03	-0.01	0.01	0.00	-0.07	-0.03	0.05
	(0.10)	(0.11)	(0.08)	(0.09)	(0.10)	(0.09)	(0.20)	(0.22)
Partial × Written with GPT	-0.08	0.03	-0.09	-0.01	-0.06	0.01	-0.03	-0.12
	(0.15)	(0.15)	(0.13)	(0.14)	(0.15)	(0.14)	(0.29)	(0.32)
Full × Written with GPT	-0.01	-0.03	-0.13	0.02	-0.11	-0.08	-0.22	-0.31
	(0.14)	(0.15)	(0.13)	(0.13)	(0.14)	(0.13)	(0.27)	(0.30)
t-test Partial: ChatGPT vs. No	0.28	0.64	0.32	0.99	0.59	0.62	0.87	0.77
t-test Full: ChatGPT vs. No	0.66	0.92	0.19	0.76	0.28	0.14	0.30	0.26
t-test Partial vs. Full: No ChatGPT	0.77	0.47	0.21	0.36	0.39	0.98	0.16	0.07
t-test Partial vs. Full: Yes ChatGPT	0.92	0.81	0.28	0.41	0.39	0.62	0.56	0.38
Control Group Mean	-0.06	-0.02	-0.05	-0.02	-0.01	-0.02	3.36	0.52
Control Group S.D.	1.02	0.98	0.99	1.02	0.99	0.98	1.20	0.50
Observations	830	830	830	830	830	830	830	830

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses, and are clustered at the recruiter-level. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Sample consists of only Male recruiters. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A22: Old Recruiters: Effect of Knowledge About ChatGPT Usage on Cover Letter Evaluation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter			Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
Partial	-0.03	-0.12	-0.09	-0.06	-0.04	-0.09	-0.18	-0.29
	(0.13)	(0.14)	(0.14)	(0.13)	(0.14)	(0.14)	(0.21)	(0.21)
Full	0.14	0.12	0.13	0.14	0.15	0.18	0.20	0.14
	(0.15)	(0.15)	(0.15)	(0.15)	(0.16)	(0.14)	(0.23)	(0.22)
Cover Letter Written with GPT	-0.04	0.10	0.01	-0.01	0.01	-0.07	-0.15	-0.18
	(0.10)	(0.12)	(0.11)	(0.09)	(0.10)	(0.10)	(0.20)	(0.21)
Partial × Written with GPT	0.05	0.05	-0.01	0.04	0.06	0.10	0.16	0.20
	(0.15)	(0.15)	(0.15)	(0.13)	(0.14)	(0.14)	(0.26)	(0.30)
Full × Written with GPT	0.07	0.02	0.05	0.10	-0.00	0.03	0.10	0.18
	(0.16)	(0.16)	(0.16)	(0.15)	(0.16)	(0.16)	(0.29)	(0.32)
t-test Partial: ChatGPT vs. No	0.99	0.20	0.98	0.84	0.55	0.84	0.92	0.90
t-test Full: ChatGPT vs. No	0.83	0.33	0.60	0.46	0.92	0.72	0.84	0.99
t-test Partial vs. Full: No ChatGPT	0.02	0.00	0.00	0.00	0.04	0.00	0.01	0.01
t-test Partial vs. Full: Yes ChatGPT	0.07	0.05	0.01	0.02	0.20	0.06	0.08	0.08
Control Group Mean	0.01	0.06	0.06	-0.00	0.03	0.02	3.40	0.54
Control Group S.D.	1.06	1.02	1.05	1.03	1.02	1.05	1.26	0.50
Observations	945	945	945	945	945	945	945	945

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses, and are clustered at the recruiter-level. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Sample consists of only recruiters above the median age. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Table A23: Young Recruiters: Effect of Knowledge About ChatGPT Usage on Cover Letter Evaluation

	(1) Total	(2) Layout	(3) Intro	(4) Cover Letter Experience	(5) Motivation	(6) Closing	(7) Likelihood of Interview	(8) High Chance of Interview
Partial	-0.01	-0.04	0.09	0.04	-0.01	-0.01	-0.09	-0.06
	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)	(0.11)	(0.22)	(0.23)
Full	0.02	0.07	0.14	0.08	0.06	0.05	0.21	0.35*
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.21)	(0.21)
Cover Letter Written with GPT	-0.03	-0.03	0.00	0.02	-0.03	-0.05	-0.04	-0.01
	(0.08)	(0.09)	(0.08)	(0.09)	(0.09)	(0.08)	(0.18)	(0.20)
Partial × Written with GPT	0.03	0.25*	0.08	0.01	0.05	0.03	0.12	0.09
	(0.13)	(0.13)	(0.14)	(0.13)	(0.13)	(0.13)	(0.26)	(0.30)
Full × Written with GPT	-0.03	0.03	-0.11	-0.05	-0.13	-0.05	-0.40*	-0.59**
	(0.12)	(0.12)	(0.11)	(0.12)	(0.12)	(0.11)	(0.24)	(0.25)
t-test Partial: ChatGPT vs. No	0.97	0.04	0.41	0.80	0.87	0.82	0.49	0.72
t-test Full: ChatGPT vs. No	0.52	0.97	0.26	0.70	0.09	0.27	0.05	0.00
t-test Partial vs. Full: No ChatGPT	0.99	0.68	0.57	0.76	0.98	0.76	1.00	0.57
t-test Partial vs. Full: Yes ChatGPT	0.77	0.41	0.16	0.90	0.34	0.84	0.15	0.23
Control Group Mean	-0.01	-0.06	-0.06	0.00	-0.03	-0.02	3.40	0.52
Control Group S.D.	0.94	0.98	0.95	0.97	0.98	0.95	1.15	0.50
N	1060	1060	1060	1060	1060	1060	1060	1060

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses, and are clustered at the recruiter-level. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Sample consists of only recruiters below the median age. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A.2.4 Partial Information Only - Perceptions of Recruiters

In this section, we replicate Table 6 among recruiters in the *Partial Information* treatment, with the independent variable being the recruiter's belief of whether the cover letter was written with or without the assistance of LLMs. While the independent variable is not exogenous and hence causality cannot be claimed, the results provide suggestive evidence that recruiters evaluate cover letters that they perceive were written with the use of LLMs more positively, however this does not translate into a higher interview likelihood.

Table A24: Partial Info Only: Effect of Perception About ChatGPT Usage on Cover Letter Evaluation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter			Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
Thinks CL Written With GPT	0.15*	0.28***	0.20**	0.07	0.05	0.18**	0.19	0.16
	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)	(0.16)	(0.18)
Control Group Mean	-0.07	-0.14	-0.08	-0.03	-0.04	-0.10	3.32	0.49
Control Group S.D.	0.99	0.95	0.99	0.92	0.99	0.97	1.16	0.50
Observations	660	660	660	660	660	660	660	660

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (when recruiters thought the cover letter was not written with LLM assistance). Robust standard errors are in parentheses. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Panel A reports treatment effects for all cover letters, while Panels B and C report treatment effect for cover letters written without LLM, and with LLM assistance, respectively. Sample consists of only recruiters in the *Partial Information* treatment. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A.2.5 Tercile Heterogeneity

Table A25 reports treatment effects for cover letters in the lower tercile. The evaluations of the recruiters in the recruiter-side experiment are similar to those of recruiters in the job-seeker Experiment, as the average evaluation of these cover letters is below the mean, indicated by the negative value of the standardized control group mean (referring to the *Partial Info* treatment). Cover letters written with the assistance of ChatGPT are scored higher than those without (comparing the Control Group Means in Panels B vs. C), in line with the findings from Figure 1 which illustrates that the positive treatment effects of ChatGPT assistance on a cover letter's quality are primarily driven by lower-quality applicants. Furthermore, complete information on whether the applicant used ChatGPT or not does not impact either the evaluation of the quality of the cover letter (Columns 1 - 6), nor the likelihood of inviting the candidate to a job interview (Columns 7-8). Therefore, revealing the LLM assistance status of the applicant does not have an effect on the recruiters' evaluations of the low-quality cover letters.

Table A25: Main Results: Effect of Knowledge About ChatGPT Usage on Cover Letter Evaluation - Lower Tercile

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter			Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
Partial	0.13	0.07	0.03	0.09	0.12	0.04	0.28	0.19
	(0.15)	(0.15)	(0.14)	(0.15)	(0.15)	(0.15)	(0.27)	(0.29)
Full	0.05	0.06	0.07	0.07	0.03	0.07	0.24	0.20
	(0.14)	(0.14)	(0.13)	(0.13)	(0.14)	(0.14)	(0.25)	(0.29)
Cover Letter Written with GPT	0.44***	0.40***	0.34***	0.38***	0.38***	0.48***	0.86***	0.83***
	(0.12)	(0.13)	(0.12)	(0.11)	(0.12)	(0.13)	(0.23)	(0.25)
Partial × Written with GPT	-0.18	-0.09	-0.06	-0.09	-0.23	-0.14	-0.42	-0.38
	(0.18)	(0.18)	(0.18)	(0.17)	(0.18)	(0.18)	(0.33)	(0.38)
Full × Written with GPT	-0.01	-0.00	0.02	0.03	-0.10	-0.04	-0.28	-0.29
	(0.17)	(0.17)	(0.16)	(0.16)	(0.17)	(0.17)	(0.31)	(0.36)
t-test Partial: ChatGPT vs. No	0.04	0.02	0.03	0.02	0.22	0.01	0.06	0.11
t-test Full: ChatGPT vs. No	0.00	0.00	0.00	0.00	0.04	0.00	0.01	0.05
t-test Partial vs. Full: No ChatGPT	1.00	0.67	0.36	0.72	0.74	0.47	0.90	0.76
t-test Partial vs. Full: Yes ChatGPT	0.49	0.51	0.29	0.48	0.84	0.35	0.62	0.69
Control Group Mean	-0.40	-0.39	-0.30	-0.36	-0.35	-0.34	2.92	0.38
Control Group S.D.	1.05	1.04	1.01	1.00	1.04	1.05	1.25	0.49
N	669	669	669	669	669	669	669	669

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses, and are clustered at the recruiter-level. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Sample consists of cover letters that were identified as being part of the lower tercile, based on evaluations from the first Experiment. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

Focusing next on the medium-quality cover letters - those which were graded in the middle tercile of the job-seeker experiment - Table A26 highlights heterogeneity in the evaluations of cover letters written with and without the assistance of LLMs. The perceived quality of the cover letter is unchanged as a result of *Full Information*, as the average grade of the cover letter is the same for both types of cover letters (see the Control Group Mean in Panels B and C), and the treatment effects are indistinguishable in Columns 1-6 across Panels A-C. Despite evaluating the cover letters as equally good, recruiters were statistically significantly more likely to recommend the applicant to the next step of the application process. This can be seen by comparing Columns 7 and 8 across Panels B and C. Informing recruiters that a cover letter was written without the assistance of ChatGPT statistically significantly increased the likelihood and chance of inviting the applicant to a job interview. Similarly, informing the recruiter that a cover letter was written with the assistance of ChatGPT reduced the likelihood of inviting the applicant to a job interview, albeit not statistically significantly. This suggests that recruiters place a premium on applications that did not use ChatGPT, even if the cover letters are deemed to be of comparable, average quality.

Table A26: Main Results: Effect of Knowledge About ChatGPT Usage on Cover Letter Evaluation - Medium Tercile

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter			Likelihood of	High Chance of
	Total	Layout	Intro	Experience	Motivation	Closing	Interview	Interview
Partial	-0.07	-0.12	-0.01	-0.07	-0.05	0.01	-0.31	-0.19
	(0.13)	(0.13)	(0.14)	(0.13)	(0.13)	(0.12)	(0.23)	(0.27)
Full	0.03	0.05	0.10	0.06	0.16	0.15	0.18	0.40
	(0.13)	(0.13)	(0.14)	(0.13)	(0.13)	(0.12)	(0.24)	(0.28)
Cover Letter Written with GPT	-0.09	-0.02	-0.10	-0.10	0.06	0.01	-0.22	-0.21
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.23)	(0.26)
Partial × Written with GPT	0.11	0.24	0.10	0.12	0.08	-0.02	0.39	0.33
	(0.16)	(0.16)	(0.17)	(0.16)	(0.16)	(0.16)	(0.33)	(0.38)
Full × Written with GPT	0.05	0.03	-0.02	0.07	-0.22	-0.09	-0.17	-0.47
	(0.16)	(0.16)	(0.17)	(0.16)	(0.15)	(0.15)	(0.34)	(0.39)
t-test Partial: ChatGPT vs. No	0.90	0.07	1.00	0.88	0.24	0.92	0.57	0.66
t-test Full: ChatGPT vs. No	0.75	0.90	0.37	0.80	0.17	0.50	0.11	0.01
t-test Partial vs. Full: No ChatGPT	0.42	0.48	0.57	0.21	0.42	0.16	0.20	0.37
t-test Partial vs. Full: Yes ChatGPT	0.73	0.73	0.97	0.51	0.50	0.48	0.87	0.45
Control Group Mean	0.22	0.21	0.17	0.16	0.23	0.23	3.62	0.59
Control Group S.D.	0.93	0.91	0.93	0.97	0.92	0.92	1.14	0.49
N	668	668	668	668	668	668	668	668

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses, and are clustered at the recruiter-level. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). Sample consists of cover letters that were identified as being part of the medium tercile, based on evaluations from the first Experiment. ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

A.2.6 No Clustered Standard Errors

Table A27: Main Results: Effect of Knowledge About ChatGPT Usage on Cover Letter Evaluation

	(1) Total	(2) Layout	(3) Intro	(4) Cover Letter Experience	(5) Motivation	(6) Closing	(7) Likelihood of Interview	(8) High Chance of Interview
Partial	-0.03	-0.10	-0.01	-0.02	-0.04	-0.05	-0.14	-0.16
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.15)	(0.16)
Full	0.06	0.08	0.12	0.10	0.09	0.10	0.19	0.26*
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.15)	(0.16)
Cover Letter Written with GPT	-0.04	0.03	0.01	0.00	-0.01	-0.06	-0.10	-0.09
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.15)	(0.16)
Partial × Written with GPT	0.04	0.14	0.03	0.02	0.05	0.06	0.14	0.14
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.20)	(0.22)
Full × Written with GPT	0.02	0.02	-0.04	0.02	-0.07	-0.01	-0.15	-0.25
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.20)	(0.22)
t-test Partial: ChatGPT vs. No	0.99	0.02	0.64	0.78	0.57	0.99	0.61	0.76
t-test Full: ChatGPT vs. No	0.82	0.45	0.71	0.77	0.29	0.34	0.12	0.03
t-test Partial vs. Full: No ChatGPT	0.12	0.03	0.08	0.03	0.19	0.03	0.08	0.03
t-test Partial vs. Full: Yes ChatGPT	0.31	0.41	0.40	0.11	0.92	0.28	0.82	0.80
Control Group Mean	0.00	0.00	0.00	0.00	0.00	0.00	3.40	0.53
Control Group S.D.	1.00	1.00	1.00	1.00	1.00	1.00	1.20	0.50
N	2005	2005	2005	2005	2005	2005	2005	2005

Notes: Intention to Treat estimates. Column 1-6 report treatment effects from OLS regressions, while columns 7 and 8 report multinomial logit and logit regressions, respectively. All regressions include controls for the recruiter's age and sex, as well as job-applicant fixed effects. Standard errors are clustered at the recruiter level. Control mean refers to the mean value of the outcome in the control group (No Information). Robust standard errors are in parentheses, and are clustered at the recruiter-level. Column 1-7 refer to variables as described in Appendix B.1.3, while Column 8 refers to a dummy if the Likelihood of Being Invited to an Interview is greater than three (on a five-point scale). ***, ** and * represent significant differences at the 1, 5 and 10% level, respectively.

B Experiment Logistics

B.1 Job-seeker Experiment

B.1.1 Blocked Websites

OpenAI	Perplexity	Gemini
Falcon LLM	Google Gemini	Huggingface
Mistral	Llama	Claude
Anthropic	CopyAI	Anyword
Sudowrite	Writer	Writesonic
Rytr	Jasper AI	Simplified
Wordai	Grammarly	Careerflow AI
Resume IO	Kickresume	Teal HQ
Google Drive	Google Mail	Dropbox
Onedrive	SurfDrive	Rezi AI Cover Letter Builder
Jobscan	Coverletter Copilot	Microsoft Copilot
Myperfectresume	Coverdoc AI	AI Apply
Lazyapply	Zety	Aicoverlettergenerator
Coverletter-ai	Easycoverletter	Master Interview AI
There's an AI for that		

Table A28: List of Blocked Domain Names

OpenAI, in bold, is the only domain page that is blocked in the Control group, but not in the Treatment group.

B.1.2 Training Materials

The training material for the Control group can be found here.
The training material for the Treatment group can be found here.

B.1.3 Evaluation Criteria

	Low (0-4)	Medium (4-7)	High (7-10)
Layout, Writing Quality, and Clarity	Not professional in appearance; poor formatting; No flow or order to the way things are discussed; spelling and grammar errors; confusing sentences or main points	Letter generally looks clear and professional; Generally able to follow organization and flow; very few mistakes in spelling and grammar; some connection between the narrative and the position, but not maximized	Professional appearance; clean fonts and formatting; Clear organization; clean and consistent layout; free of grammar, spelling errors; overall narrative that clearly connects main points
Introduction	Does not cover basic info; weak link to the body of the cover letter; no link between the applicant and the position.	Conveys basic information in an unengaging form. Links to the main part of the cover letter, but not very effectively.	Covers basic information, offers an engaging and gripping way into the body of the letter and clearly connects the person to the position.
Relevant Experiences and Demonstration of Skills	not enough/ too much information in key areas; background, education and experience not fully explained; more questions raised than answered; all assertions without foundation or specifics to support them.	background, education and experience laid out but not connected to the position; Vague but inconsistent narrative.	Clear narrative; outlines background, education and experience fully and with specifics that connect directly to the position.
Motivation for Role and Alignment with Job	Little discussion of their motivation, the company, or how they would fit into existing organization.	Decent, but not always coherent motivation for the role; some good connections between the position and the applicant; some research done about the firm, but not that much.	focus on how they fit the job and will be effective members of the organization; outlines motivation in a coherent manner; strong understanding of the role, the organization, and how they would fit within the organization.
Closing	Not a strong closing statement; repetitive or wandering; no clear 'final message' to the reader and how they fit the job as outlined.	Has a sense of a closing statement but unenthusiastic or unconvincing; tries to convey too much, making it confusing; no succinct final message.	Strong closing statement of purpose; clearly outlines how their background, education and experience have prepared them for this specific position (without being repetitive).

Figure A8. Cover Letter Evaluation Criteria

	Low (0-4)	Medium (4-7)	High (7-10)
Layout, Writing Quality, and Clarity	Not professional in appearance; poor formatting; No flow or order to the way things are discussed; spelling and grammar errors; confusing sentences or main points.	CV generally looks clear and professional; Generally able to follow organization and flow; very few mistakes in spelling and grammar; descriptions are generally clear.	Professional appearance; clean fonts and formatting; Clear organization; clean and consistent layout; free of grammar, spelling errors; effective use of CV phrasing.
Education	Below-average student, or from a subject not related to position they are applying for.	Average student, however does not belong to the top share of their cohort.	One of the top students in their cohort, as part of a demanding and relevant degree.
Experience	No or little relevant work / internship, or leadership experience.	Some relevant work / internship, or leadership experience, however it could be more.	Strong relevant work / internship, or leadership experience, for example through work experience or student associations.
Extra-curricular	No evidence of involvement in extra-curricular activities.	Some involvement in extra-curricular activities.	Strong involvement in extra-curricular activities.

Figure A9. CV Evaluation Criteria

B.2 Recruiter Experiment

Figure A10. Cover Page

(a) No Info

In the spring of 2024, university students wrote a cover letter for a Traineeship at Rabobank, one of the largest banks in the Netherlands.

You will be asked to **evaluate FIVE cover letters**, based on a pre-defined grading rubric. Please grade them as if you were a recruiter of the company the students are applying to.

The job description is the following:

(b) Partial Info

In the spring of 2024, university students wrote a cover letter for a Traineeship at Rabobank, one of the largest banks in the Netherlands.

You will be asked to **evaluate FIVE cover letters**, based on a pre-defined grading rubric. Please grade them as if you were a recruiter of the company the students are applying to.

Some of the Cover Letters were written with the assistance of ChatGPT, while others were NOT.

You will be asked to indicate whether you think the Cover Letter was written using the help of ChatGPT, or not.

The job description is the following:

(c) Full Info

In the spring of 2024, university students wrote a cover letter for a Traineeship at Rabobank, one of the largest banks in the Netherlands.

You will be asked to **evaluate FIVE cover letters**, based on a pre-defined grading rubric. Please grade them as if you were a recruiter of the company the students are applying to.

Some of the Cover Letters were written with the assistance of ChatGPT, while others were NOT.

We WILL indicate whether the Cover Letter was written using the help of ChatGPT, or not, before you read the cover letter.

The job description is the following:

Figure A11. Cover Letter Page

(a) No Info

Below you will find a cover letter, written for the job description you just saw, which you can find here.

You will have to evaluate the cover letter, based on the evaluation rubric, which you can find here.

(b) Partial Info

Below you will find a cover letter, written for the job description you just saw, which you can find here.

You will have to evaluate the cover letter, based on the evaluation rubric, which you can find here.

(c) Full Info

Below you will find a cover letter, written for the job description you just saw, which you can find here.

While writing the cover letter, **the student WAS NOT allowed to use ChatGPT.**

You will have to evaluate the cover letter, based on the evaluation rubric, which you can find here.

C Recruiter Insights on the Labor Market

The recruiter-side survey was also used to gain further insights into recruiters' perceptions of LLMs, and cover letters. These are outlined below:

Recruiters were asked to evaluate the sub-components of the cover letter (Layout, Introduction, Experience, Motivation, Conclusion) on a scale of 1-5, where 1 indicated the most personalized part of the cover letter, while 5 indicated the least personalized part of the cover letter.

Table A29: Personalization of Cover Letter Components

Cover Letter Component	Degree of Personalization
Layout	3.72
Introduction	2.46
Experience	2.44
Motivation	2.24
Conclusion	4.15

This indicates that the most personalized sections of the cover letter are the Motivation and Introduction. The least personalized section is the Conclusion.

Recruiters were also asked, on a five-point Likert scale, the degree to which they agree or disagree with the following statements: "It is acceptable for job applicants to use LLMs in their application"; "Job applicants should disclose LLM usage in their applications"; "The use of LLMs (e.g., ChatGPT) in a cover letter improves their grammar"; "The use of LLMs (e.g., ChatGPT) in a cover letter increases its originality and personalization". The results are presented in the table below:

Table A30: Agree-ability with Statements

	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree	Average Value
Acceptable to Use	7.48%	24.19%	18.95%	40.40%	8.98%	3.19
Should Disclose Usage	8.23%	16.46%	26.1%8	27.68%	21.45%	3.38
LLM improves Grammar	2.74%	11.22%	17.21%	48.13%	20.70%	3.73
LLM increases originality	24.19%	38.40%	20.45%	13.72%	3.24%	2.33

Recruiters were asked on a five-point Likert scale how confident they were that they could detect LLM usage by job applicants. The results are presented in the Table below:

Table A31: Recruiter Confidence Detecting Use of LLMs in Cover Letters

	Not	Slightly	Moderately Confident	Very	Extremely	Average Value
Confident Detect LLM Use	5.51%	34.09%	42.61%	15.54%	2.26%	2.75

Recruiters were also asked about their attitudes towards applicants using LLMs in their cover letters:

Table A32: Recruiter perceptions about Use of LLMs in Cover Letters

	Very Negative	Negative	Neutral	Positive	Very Positive	Average Value
LLM Use in Cover Letter	7.29%	23.87%	44.72%	21.86%	2.26%	2.88

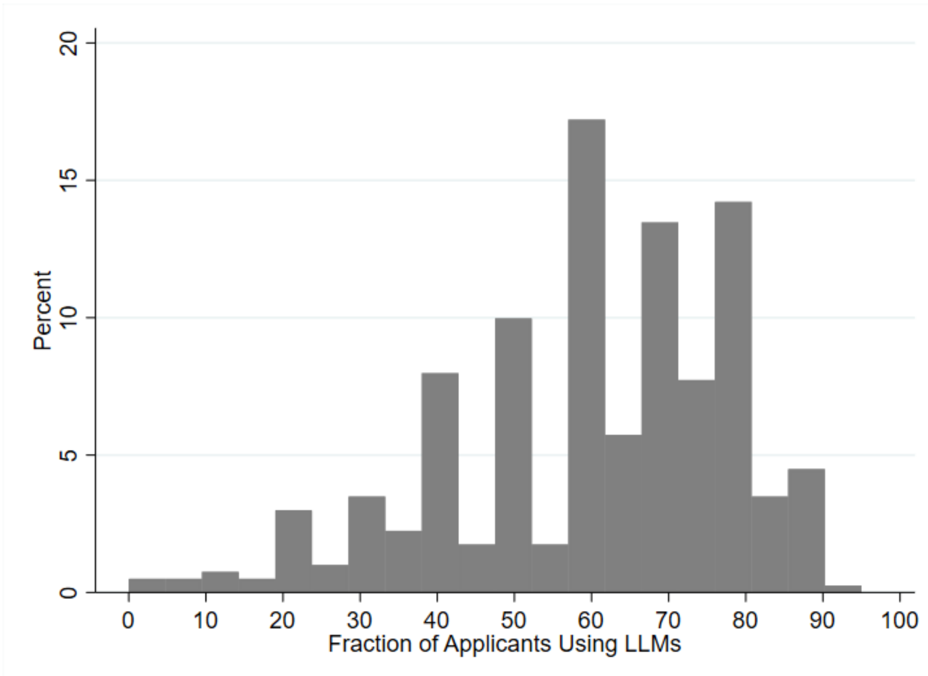
Recruiters were asked whether they thought LLMs like ChatGPT would have a smaller or greater impact on the quality of job applications, compared with Algorithmic Writing Assistants like Grammarly, which were empirically evaluated by Wiles et al. (2025).

Table A33: Impact of LLMs vs. Algorithmic Writing Assistants

	Far Smaller	Somewhat Smaller	Equally-sized	Somewhat Larger	Far Larger	Average Value
Impact of LLMs vs. AWA	3.99%	7.98%	19.20%	48.13%	20.70%	3.74

Lastly, recruiters were asked what percentage of applicants they think currently use LLMs in their job applications. The mean was 60.63%, with a wide distribution as illustrated by Figure A12.

Figure A12. Histogram of Recruiter's Perception on LLM Use Among Applicants



D ChatGPT Conversation Analysis

This section details our methodology for analyzing conversations between users and ChatGPT during cover letter writing. We use OpenAI API with model *gpt-4o-mini* to classify and quantify user interactions, enabling nuanced analysis beyond simple keyword matching.

D.1 Methodology

We analyze six key cover letter components, classifying user messages into three categories:

- **None**: Message does not pertain to the specific section.
- **Content**: User seeks substantive content related to the section.
- **Guidance**: User requests structural/formatting advice.

Our analysis covers these sections:

Layout: Formatting, structure, presentation, grammar, readability
Introduction: Opening statements, salutations, position identification
Experience: Professional background, education, skills, qualifications
Motivation: Personal drive, interest in position/company, career goals
Conclusion: Closing statements, next steps, gratitude, contact information
Generic: Broad requests like ”write me a cover letter” without section-specific focus

Additionally, we analyze four meta-categories:

Company Research: Questions about company info, values, culture, industry
General Strategy: Overall approach, structure planning, best practices
Formatting: Length, font, spacing, visual presentation
Revision Requests: Improvements, editing, refining, changes

D.2 API Implementation

We developed a Python script that interfaces with the OpenAI API. The system prompt instructs:

```
{
    "role": "system",
    "content": "Analyze this conversation between a user and ChatGPT about cover
    ↪  letter writing.
```

```

For each section below, classify the user's engagement:

SECTIONS:
- layout: Formatting, structure, presentation, grammar, readability
- introduction: Opening statements, salutations, position identification
- experience: Professional background, education, skills, qualifications
- motivation: Personal drive, interest in position/company, career goals
- conclusion: Closing statements, next steps, gratitude, contact info
- generic: Broad requests like \"write me a cover letter\", general help without
  ↪ specific section focus

For each section, classify as:
- \"none\": No engagement with this section
- \"guidance\": User asks for advice/tips on how to approach the section
- \"content\": User asks ChatGPT to generate or improve actual text for the
  ↪ section

Return ONLY a JSON object:
{
  \"layout\": \"none|guidance|content\",
  \"introduction\": \"none|guidance|content\",
  \"experience\": \"none|guidance|content\",
  \"motivation\": \"none|guidance|content\",
  \"conclusion\": \"none|guidance|content\",
  \"generic\": \"none|guidance|content\"
}

Base analysis only on USER messages."
}
```

```
"role": "system",
    "content": "Analyze if this user engaged with these meta-categories during
    ↪ cover letter writing:
```

```

CATEGORIES:
- company_research: Questions about company info, values, culture, industry
- general_strategy: Overall approach, structure planning, best practices
- formatting: Length, font, spacing, visual presentation
- revision_requests: Improvements, editing, refining, changes

Return ONLY a JSON object:
{
  \"company_research\": true|false,
  \"general_strategy\": true|false,
  \"formatting\": true|false,
  \"revision_requests\": true|false"
}
```

D.3 Additional Analysis

D.3.1 Meta-Category Engagement

Figure A13. Meta-Category Engagement

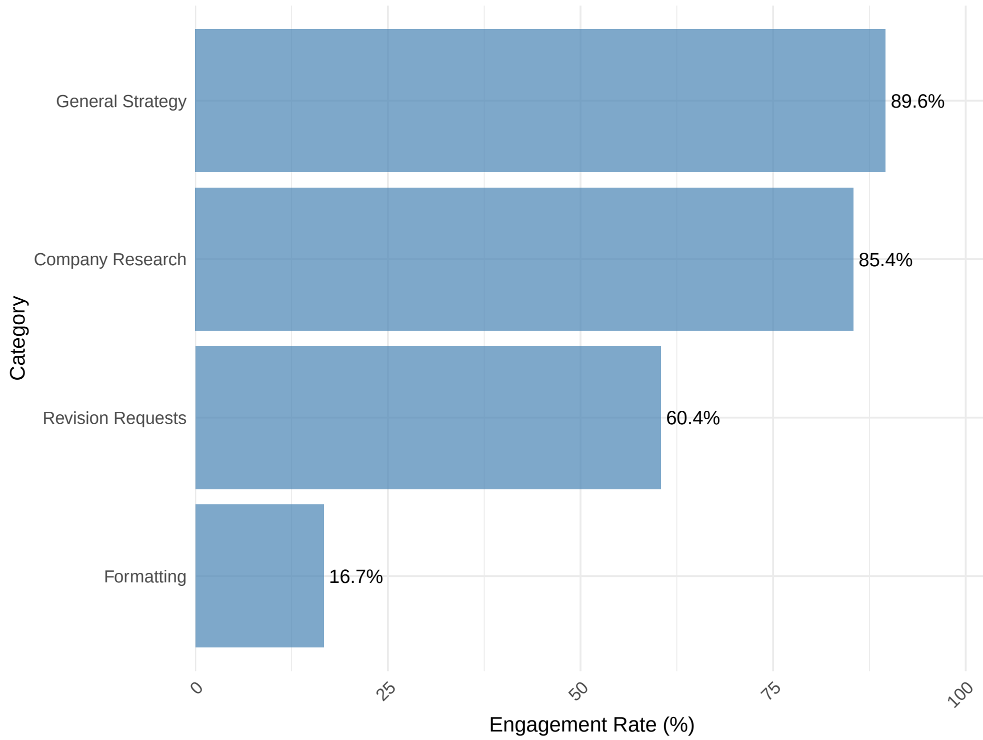


Figure A13 displays the percentage of students who engaged with each meta-category during their ChatGPT interactions for cover letter writing.

D.3.2 Message Distribution

Table A34: Summary Statistics for Number of User Messages

Agent	Mean	Median	Variance	SD	Min	Max
User	6.416667	5	19.22695	4.384855	1	16

Figure A14. Distribution of User Messages per Conversation

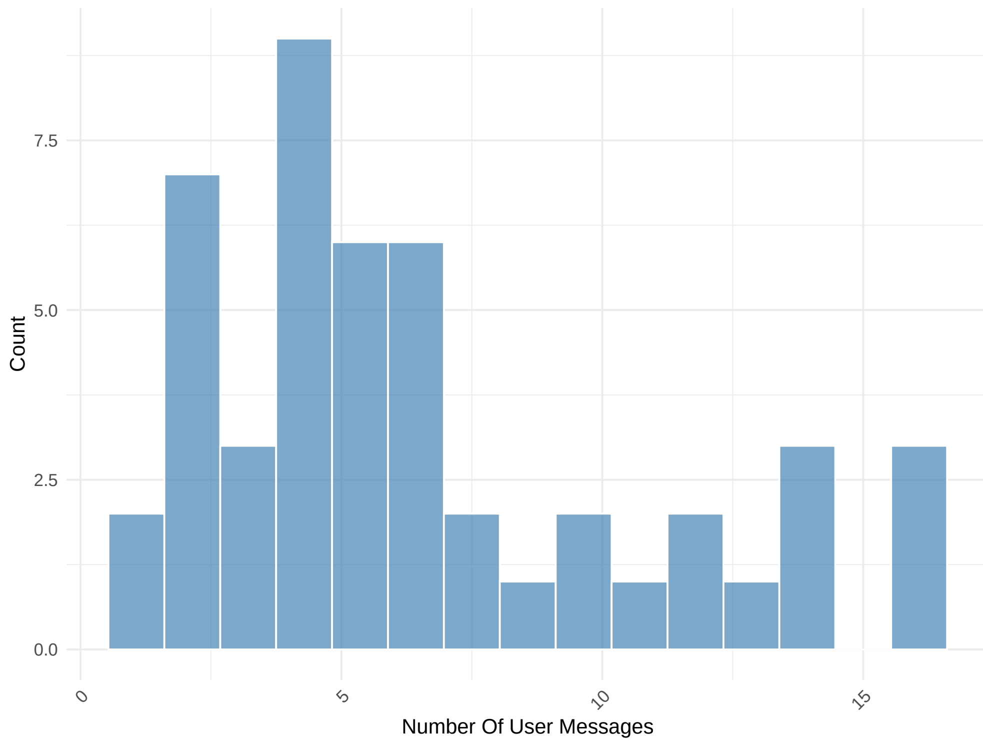


Figure A14 shows the frequency distribution of messages exchanged between users and ChatGPT, demonstrating substantial variation in user engagement patterns.

E Model Details, Proofs and Extension

This section provides further details on the model. In addition to providing a proof of proposition 1 in the main text; we provide proofs of some properties of assignment models in the case of a continuum of agents. Many of these properties—such as uniqueness of assortative matching and optimality of positive assortative matching with supermodularity—are well-known in the case of discrete number of agents. We provide details on the extentions of these results to the case of continuum of agents; to our knowledge we are the first to do so.

E.1 Perfect Information Economy

The static economy consists of a continuum of workers, indexed by their quality s and distributed with CDF G_s, and a continuum of firms, indexed by their quality x and distributed with CDF G_x. We assume that both G_s and G_x admit densities denoted by g_x and g_s and have positive and bounded supports $\mathcal{S} := [\underline{s}, \overline{s}]$ and $\mathcal{X} := [\underline{x}, \overline{x}]$. The value of a match between a worker of type s and a firm of type x is given by $f(x, s)$ which satisfies Assumption 1:

Assignment in the economy without information frictions is defined by a function $x = \sigma(s)$, which matches each worker s with a firm x. In equilibrium, all firms and workers must be matched with each other, and matches are one-to-one: as explained above, there are no multi-worker firms or multi-firm workers. Therefore, the key property of σ is that, when considered as a set transformation, it must match an equal mass of workers and firms, in other words, σ must be a *measure-preserving* transformation. Given an assignment function σ, the aggregate value (or welfare) in this economy with perfect information is given by

$$V_P = \int_{\mathcal{S}} f(\sigma(s), s) \mathrm{d}G_s(s) \tag{10}$$

The monotonically-increasing assignment function that assigns top firms with top workers and bottom firms with bottom workers is known as *positively assortative matching* (PAM). PAM essentially matches the top $q = G_x(x)$ quantile of firms with the top $q = G_s(s)$ quantile of workers, and thus is given by $x = \sigma_P(s) = G_x^{-1}(G_s(s))$. Lemma 1 shows that PAM is unique in that it is the only monotonically-increasing measure-preserving transformation from the set of workers to the set of firms.

Lemma 1. *PAM is the only monotonically-increasing assignment function.*

Proof of Lemma 1. The relevant probability spaces are $(\mathcal{S}, \mathcal{B}_s, P_s)$ and $(\mathcal{X}, \mathcal{B}_x, P_x)$ where $\mathcal{B}_s, \mathcal{B}_x$ are the Borel σ-algebras of $\mathcal{S}, \mathcal{X}$, and P_s,P_x are the probability measures associated with distribution functions G_s, G_x.

An assignment function, as explained in the main text, is a function $\sigma : \mathcal{S} \to \mathcal{X}$ that assigns each worker $s \in \mathcal{S}$ to a firm $x \in \mathcal{X}$. Considered as a set transformation, it is a mapping from $\mathcal{B}_s$ to $\mathcal{B}_x$ that assigns groups of workers $S \subset \mathcal{S}$ to groups of firms $X \subset \mathcal{X}$. In equilibrium, an assignment transformation must be measure-preserving, i.e.

$$P_s\big\{\sigma^{-1}(B)\big\} = P_x\{B\} \; \forall B \in \mathcal{B}_x \tag{11}$$

Consider the sets $B = (\underline{x}, t)$ of the Borel σ-algebra $\mathcal{B}_x$, for any $\underline{x} < t \le \overline{x}$. If σ is monotonically increasing, i.e. $\sigma(t') > \sigma(t) \; \forall t' > t$, this implies that:

$$P_s\big\{\sigma^{-1}((\underline{x}, t))\big\} = P_s\big\{(\underline{s}, \sigma^{-1}(t))\big\} = G_s(\sigma^{-1}(t)). \tag{12}$$

Where we used the fact that $\sigma^{-1}(\underline{x}) = \underline{s}$, since otherwise $\underline{x}$ or $\underline{s}$ will be unmatched, violating preservation of measure. Then Equation (11) gives:

$$G_s(\sigma^{-1}(t)) = G_x(t)$$

implying that σ must be PAM: $\sigma(t) = G_x^{-1}(G_s(t))$.

When dealing with monotonically-decreasing assignments, Equation (12) becomes

$$P_s\big\{\sigma^{-1}((\underline{x}, t))\big\} = P_s\big\{(\sigma^{-1}(t), \overline{s})\big\} = 1 - G_s(\sigma^{-1}(t)) \tag{13}$$

which leads to *negative assortative matching*: $G_x^{-1}(1 - G_s(t))$ being the unique monotonically-decreasing assignment. ■

In the context of discrete distributions Becker (1973) has shown that under the assumption of complementarity, PAM is the optimal assignment that maximizes Equation (10). We briefly extend this to the case of a continuum of workers and firms.

Lemma 2. *The positive assortative matching $\sigma_P(s) = G_x^{-1}(G_s(s))$ is optimal.*

Proof of Lemma 2. The relevant probability space is $(\mathcal{S}, \mathcal{B}_s, P_s)$ where $\mathcal{B}_s$ is the Borel σ-algebra of $\mathcal{S}$, and P_s is the probability measure associated with distribution function G_s.

For this proof we use the rearrangement inequality theorem of Burchard and Hajaiej (2006), who extend the results of Crowe et al. (1986), Almgren and Lieb (1989), and Brock (2000). This theorem, a generalization of the Hardy-Littlewood inequality, states that if f is a supermodular function, the following inequality holds for any measurable functions u, v:

$$\int_{\mathcal{S}} f(u(s), v(s))\mathrm{d}P_s \leq \int_0^1 f(\overline{u}(z), \overline{v}(z))\mathrm{d}z \tag{14}$$

where $\overline{u}(s)$ is the unique non-increasing rearrangement of $u(s)$ defined as:

$$\overline{u}(z) := \sup\{t \geq 0 : \rho_u(t) \geq z\}$$

where $\rho_u(t) = P_s(\{s \in \mathcal{S} : u(s) > t\})$ is the distribution function of u (a similar definition holds for $\overline{v}(s)$).

Intuitively, a non-increasing rearrangement of a function u is function $\overline{u}$ whose level sets have the same measure as u but are re-arranged in a decreasing order. For more information on rearrangements and related inequalities see Burchard (2009).

In our case, $u(s) = G_x^{-1}(G_s(s))$ and $v(s) = s$. Thus

$$\begin{aligned}
\rho_u(t) &= P_s(\big\{s \in \mathcal{S} : G_x^{-1}(G_s(s)) > t\big\}) = P_s(\big\{s \in \mathcal{S} : s > G_s^{-1}(G_x(t))\big\}) && (15)\\
&= 1 - G_x(t) && (16)\\
\rho_v(t) &= P_s(\{s \in \mathcal{S} : s > t\}) = 1 - G_s(t) && (17)
\end{aligned}$$

So

$$\begin{aligned}
\overline{u}(z) &= \sup\{t \geq 0 : 1 - G_x(t) \geq z\} = G_x^{-1}(1-z) && (18)\\
\overline{v}(z) &= \sup\{t \geq 0 : 1 - G_s(t) \geq z\} = G_s^{-1}(1-z) && (19)
\end{aligned}$$

so the rearrangement inequality states

$$\int_{\mathcal{S}} f(\sigma_P(s), s)\mathrm{d}G_s(s) \leq \int_0^1 f(G_x^{-1}(1-z), G_s^{-1}(1-z))\mathrm{d}z \tag{20}$$

Let I denote the rightmost integral. With a simple change of variable $z = 1 - G_s(s)$ we obtain

$$I = -\int_{\overline{s}}^{\underline{s}} f(G_x^{-1}(G_s(s)), s)\mathrm{d}G_s(s) = \int_{\underline{s}}^{\overline{s}} f(G_x^{-1}(G_s(s)), s)\mathrm{d}G_s(s) \tag{21}$$

which shows that u, v are their own nonincreasing rearrangements, thus proving the optimality of PAM. ■

E.2 Imperfect Information without LLMs

Consider now an economy where firms cannot directly observe the quality of each worker and instead observe a signal y_1

$$y_1 = s + e \tag{22}$$

where e is the signal error with mean 0, distributed with CDF G_e on support $\mathcal{E} = [\underline{e}, \overline{e}]$. Here e reflects the fact that job-seekers' signal cover letter) does not fully reflect their relevant skills and abilities. Having received a signal y_1, firms form their Bayesian estimates of the hidden worker type as $\hat{s}(y_1) = \mathbb{E}[s|y_1]$ (which has the same support $\mathcal{S}$ as the true distribution of workers).

Implicit in the signal Equation (22) is the assumption that signal values are independent of firm qualities and only depend on the worker quality. That is, we assume each worker only produces one signal, which is then observed by all firms. More intuitively this is the scenario where each worker producer one job application package and sends it to all firms (costs of job applications and posting vacancies are irrelevant to our framework).

After all firms have observed each worker's signal, they form a positive assortative matching based on estimated worker qualities $\hat{s}$. Thus the assignment of firms to estimated worker qualities is deterministic and given by $x = \sigma_I(\hat{s}) = G_x^{-1}(G_{\hat{s}}(\hat{s}))$, while the randomness in matching between firm and true worker qualities stems purely from the randomness in worker quality estimates due to the imperfect signal. Given a worker of type s, they will be matched according to their estimated quality $\hat{s}$ and so the expected output from the matches with this worker will be:

$$\mathbb{E}[f(\sigma_I(\hat{s}), s)|s] = \int_{\mathcal{S}} f(G_x^{-1}(G_{\hat{s}}(\hat{s})), s) \mathrm{d}G_{\hat{s}|s}(\hat{s}|s)$$

This implies that the total output in this economy with PAM and imperfect information is given by:

$$V_I = \int_{\mathcal{S}} \int_{\mathcal{S}} f(G_x^{-1}(G_{\hat{s}}(\hat{s})), s) \mathrm{d}G_{\hat{s}|s}(\hat{s}|s) \mathrm{d}G_s(s) \tag{23}$$

The assignment given by σ_I is positively assortative with respect to the distribution of workers' quality estimates $\hat{s}$ and by Proposition 2 it is the optimal assignment between workers and firms given the information constraints. However, as asserted by Lemma 3, aggregate value in this economy with imperfect information is at most equal to the economy with perfect information.

Lemma 3. $V_I \leq V_P$

Proof of Lemma 3. With the concavity of f, Jensen's inequality shows that the value of the economy under imperfect information (V_I) is at most equal to that of the perfect information economy

(V_P):

$$V_I \leq \int_{\mathcal{S}} f(\tilde{\sigma}(s), s)\mathrm{d}G_s(s) \leq \int_{\mathcal{S}} f(\sigma_P(s), s)\mathrm{d}G_s(s) = V_P \tag{24}$$

where $\tilde{\sigma}(s) = \mathbb{E}\big[G_x^{-1}(G_{\hat{s}}(\hat{s}))|s\big] \neq \sigma_P(s)$ for all s a.s. and the second inequality follows from Proposition 2. ■

Due to the randomness in matching between firms and true worker qualities s, some firms and workers will be better off in the scenario with imperfect information compared to the perfect information case. However there will be aggregate losses due to concavity and complementarity of f: such a random matching will assign some lower-quality firms with some higher-quality workers and vice versa, thus deviating from the first-best assignment of σ_P. Because of the complementarity, these two cannot cancel out in the aggregate, resulting in net losses.

A completely random assignment will be the case in which there is no information of the form (22), i.e. workers do not produce any CVs or cover letters to provide information about their skills to the firms. Total value of this *random* assignment economy is given by

$$V_R = \int_{\mathcal{S}} \int_{\mathcal{X}} f(x, s)\mathrm{d}G_x(x)\mathrm{d}G_s(s) \tag{25}$$

where workers and firms are assigned to each other based on their (independent) distributions. Lemma 4 shows that total value in such an economy is at most equal to the imperfect information economy with informative signals.

Lemma 4. $V_R \leq V_I$

Proof of Lemma 4. Recall V_R is given by

$$V_R = \int_{\mathcal{S}} \int_{\mathcal{X}} f(x, s)\mathrm{d}G_x(x)\mathrm{d}G_s(s)$$

Using Jensen's inequality we get

$$V_R \leq \int_{\mathcal{S}} f(x_m, s)\mathrm{d}G_s(s) =: I$$

where $x_m := \mathbb{E}[x]$ is the mean firm quality. The integral I can be simply re-written in a comparable form to V_I.

$$I = \int_{\mathcal{S}} \int_{\mathcal{S}} f(x_m, s)\mathrm{d}G_{\hat{s}|s}(\hat{s}|s)\mathrm{d}G_s(s) \tag{26}$$

In equation (26), we have the total value of an economy with the same information structure as that of V_I, but all workers are matched with the mean firm. Since σ_I is the positive assortative matching and by Proposition 2 is the optimal assignment given the joint distribution of $\hat{s}$ and s, then we must have $V_R \leq V_I$. ■

E.3 Imperfect Information with LLMs

With the details of the signal with imperfect information and LLM usage given in the main text, we can provide the proof of Proposition 1.

Proof of Proposition 1. Since $\tilde{s}$ is a function of y, total value can be rewritten as an integral with respect to the conditional distribution of signal values y

$$V_L = \int_{\mathcal{S}} \int_{\mathcal{Y}} f(G_x^{-1}(G_{\tilde{s}}(\tilde{s}(y))), s) \mathrm{d}G_{y|s}(y|s) \mathrm{d}G_s(s) \tag{27}$$

where $\mathcal{Y} = [\underline{y}, \overline{y}] = [\underline{s} + \underline{e}, \overline{s} + \overline{e}]$. The conditional distribution $G_{y|s}$ can be written, using law of total probability, as

$$G_{y|s}(x|s) = \Pr\{y \leq x|s\} = p \Pr\{s + e \leq x|s\} + (1-p) \Pr\{h(s+e) \leq x|s\} \tag{28}$$

$$= pG_{y_1|s}(x|s) + (1-p)G_{y_1|s}(h^{-1}(x)|s) \tag{29}$$

So we have the decomposition $V_L = V_{A,1} + V_{A,2}$ where

$$V_{A,1} = \int_{\mathcal{S}} \int_{\mathcal{Y}} p \cdot f(G_x^{-1}(G_{\tilde{s}}(\tilde{s}(y))), s) \mathrm{d}G_{y_1|s}(y|s) \mathrm{d}G_s(s) \tag{30}$$

$$V_{A,2} = \int_{\mathcal{S}} \int_{\mathcal{Y}} (1-p) \cdot f(G_x^{-1}(G_{\tilde{s}}(\tilde{s}(y))), s) \mathrm{d}G_{y_1|s}(h^{-1}(y)|s) \mathrm{d}G_s(s) \tag{31}$$

With a change of variables $y_1 := h^{-1}(y)$ the second integral becomes

$$V_{A,2} = \int_{\mathcal{S}} \int_{\mathcal{Y}} (1-p) \cdot f(G_x^{-1}(G_{\tilde{s}}(\tilde{s}(y))), s) g_{y_1|s}(h^{-1}(y)|s)(h^{-1})'(y) \mathrm{d}y \mathrm{d}G_s(s) \tag{32}$$

$$= \int_{\mathcal{S}} \int_{\mathcal{Y}} (1-p) \cdot f(G_x^{-1}(G_{\tilde{s}}(\tilde{s}(h(y)))), s) g_{y_1|s}(z|s)(h'(z))^{-1} h'(z) \mathrm{d}z \mathrm{d}G_s(s) \tag{33}$$

$$= \int_{\mathcal{S}} \int_{\mathcal{Y}} (1-p) \cdot f(G_x^{-1}(G_{\tilde{s}}(\tilde{s}(h(y_1)))), s) \mathrm{d}G_{y_1|s}(y_1|s) \mathrm{d}G_s(s) \tag{34}$$

By concavity of f,

$$V_L \leq \int_{\mathcal{S}} \int_{\mathcal{Y}} f(\sigma_A(y), s) \mathrm{d}G_{y_1|s}(y|s) \mathrm{d}G_s(s) =: I \tag{35}$$

Where

$$\sigma_A(y) = pG_x^{-1} G_{\tilde{s}}(\tilde{s}(y)) + (1-p) G_x^{-1} G_{\tilde{s}}(\tilde{s}(h(y)))$$

Note that V_I also can be written as

$$V_I = \int_{\mathcal{S}} \int_{\mathcal{Y}} f(\sigma_I(y), s) \mathrm{d}G_{y_1|s}(y|s) \mathrm{d}G_s(s) \tag{36}$$

where $\sigma_I(y) = G_x^{-1}(G_{\hat{s}}(\hat{s}(y)))$. Note that since the integral I in (35) is with respect to the joint distribution of y_1 and s, it provides an upper bound on the value of the economy with AI in terms of an economy without AI (but with imperfect information in terms of y_1) with a different assignment function σ_A. However since σ_I is the unique monotonically-increasing 1-to-1 assignment in such an economy and $\sigma_A \neq \sigma_I$ a.s., we have by Proposition 2 that V_I must be larger than V_L. ■

E.4 LLM Adoption Correlated With Ability

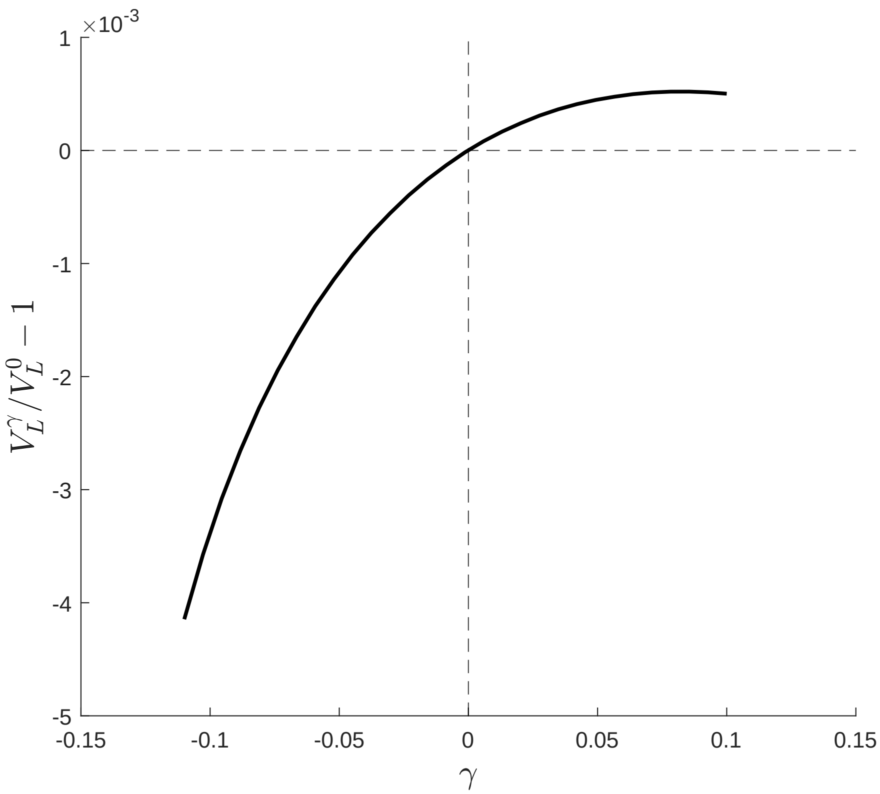
In this extension with model p as a linear function of underlying ability s, thus making the coin-flip of whether or not an applicant uses LLMs correlated with s. We write this function as:

$$p = \psi(s) := p_0 - \gamma s \tag{37}$$

Such that higher γ means *higher* correlation of underlying ability with usage of LLM (lower p). This equation implies that the average probability of adoption in the cross section is therefore equal to $p_0 - \gamma\mu_s$. In our counterfactual exercise of varying γ we fix p_0 such that this average adoption remains constant at 0.5. Therefore we set $p_0 = 0.5 + \gamma\mu_s$.

Figure A15 shows total output when varying γ between negative and positive values as compared to the baseline model where all workers use LLM with the same probability, i.e. $\gamma = 0$ and $p = p_0 = 0.5$.

Figure A15. Sensitivity of output to correlation between LLM use and skill



Notes: This figure calculates the changes in output in the economy with imperfect information and LLM technology when varying the relationship between probability of LLM adoption and worker skill, i.e. varying γ with higher γ denoting higher correlation of LLM adoption with skill. The ouotput where the use of LLM and skill are uncorrelated ($\gamma = 0$ and thus $p = 0.5$ for everyone) is chosen as the comparison benchmark.