This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Labor Market Signals: The Role of Large Language Models*

Kian Abbas Nejad[1] Giuseppe Musillo[1] Till Wicker[2] Niccolò Zaccaria[1]

[1]Tilburg University

[2]Academia Sinica

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.

*Abbas Nejad: Tilburg University (e-mail: k.abbasnejad@tilburguniversity.edu); Musillo: Tilburg University (e-mail: g.musillo@tilburguniversity.edu); Wicker: Academia Sinica (e-mail: tillwicker@econ.sinica.edu.tw); Zaccaria: Tilburg University (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 audience members at 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).

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

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, Ong, and Summers, 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, Blandin, and Deming, 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 ex-

[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.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

periment, some job-seekers were allowed to use ChatGPT 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, Li, and Raymond (2025), the positive treatment effects are stronger for lower-quality job-seekers, thus reducing the dispersion in quality of the cover letters, undermining their signaling value 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, Munyikwa, and Horton (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 eval-

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

uate 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, Munyikwa, and Horton, 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, Munyikwa, and Horton (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, Munyikwa, and Horton (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, Munyikwa, and Horton (2025) places less emphasis on soft skills, and involves shorter, resume-style documents with limited nar-

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

rative content.[2,3] 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, Munyikwa, and Horton (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, Munyikwa, and Horton (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.

Another closely related paper is Cowgill, Hernandez-Lagos, and Wright (2025), who estimate a slightly negative covariance between expertise and ChatGPT-induced signal gains through hiring and venture capital funding experiments, and show that generative AI compresses textual signals and reduces screening accuracy. Our results confirm a similar qualitative mechanism: access to LLMs improves the quality of cover letters among weaker job-seekers, compressing the

[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, Munyikwa, and Horton (2025) along several other important dimensions. Firstly, the documents evaluated differ: in Wiles, Munyikwa, and Horton (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, Munyikwa, and Horton (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, Munyikwa, and Horton (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, Munyikwa, and Horton (2025), and 2.8% in our study.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

distribution of signals and lowering recruiters' ability to screen. Our design and setting, however, yield distinct implications. Whereas evaluators in Cowgill, Hernandez-Lagos, and Wright (2025) are informed whether ChatGPT was used, we study screening under imperfectly observable LLM adoption. Recruiters are no better than chance at detecting LLM assistance, and disclosure interacts with recruiters' expectations, shifting the weight they place on high-quality human-written letters. Finally, we microfound this inference problem through firm-worker complementarity and embed it in a calibrated assignment model that maps reduced signal informativeness into welfare losses via mismatch.

In particular, findings from our second experiment shed light on how recruiters form beliefs when LLM adoption is imperfectly observable, 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, Kahn, and Li, 2017; Dell'Acqua, 2022; Avery, Leibbrandt, and Vecci, 2024; Chaturvedi and Chaturvedi, 2025; Awuah, Krenk, and Yanagizawa-Drott, 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, Metaxa, and Naaman, 2024; Bohren, Hakimov, and Lalive, 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 access to LLMs.

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

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

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 anticipate signal inflation, discount 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., 2023), consulting (Dell'Acqua et al., 2023), customer support (Brynjolfsson, Li, and Raymond, 2025), and business practices (Otis et al., 2024). 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, closing), 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 reduce readability by increasing average word length.

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

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

2 Job-seeker Experiment: Design

To study how LLM access affects job application signals, we conducted a field experiment with 137 students from Tilburg University and Utrecht University in collaboration with four large employers in the Netherlands: Philips, PwC, Rabobank, and VodafoneZiggo. Each firm provided an entry-level vacancy and two recruiters who evaluated applicants.

During a university career week, students enrolled in a *Cover Letter Challenge* by uploading a CV, providing basic demographics used for stratification, and selecting a preferred firm.[4] Participants then had one hour to write a cover letter for their chosen firm. Application packages (CV and cover letter) were pseudonymized and independently evaluated by two recruiters from the relevant firm using criteria co-developed with the firms (Appendix B.1.3). Recruiters graded multiple dimensions of the CV and cover letter, assigned overall scores, and reported the likelihood of inviting the applicant to an interview on a 5-point Likert scale. Recruiters evaluated both treated and control applications but were blind to treatment status.

Participants were informed they would receive individualized feedback and that the top three cover letters (as scored by recruiters) would receive €500, €300, and €200. Recruiters could also request to be put in touch with high-scoring applicants.

We randomized participants to a control or ChatGPT-access group. The control group viewed a short placebo guide (LinkedIn job search tips) and wrote their cover letter without access to LLM websites (Appendix B.1.1). The treatment group viewed a short guide on effective prompting and was allowed to use ChatGPT (GPT-3.5) via an account provided by the research team, while being subject to the same website restrictions otherwise.[5] Browser and conversation logs indicate high compliance: 95% of treated participants used ChatGPT, while 18% of control participants attempted (unsuccessfully) to access it. A third arm (prompting training plus ChatGPT access) was pre-registered but dropped prior to implementation due to power constraints.

Randomization was stratified by university, gender, study level (BA/MA), GPA, and age. Ap-

[4] We recruited students in the final semester of their Bachelor's or Master's program who were actively searching for a job.

[5] At the time of the experiment (Spring 2024), GPT-3.5 was the most advanced free version.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

pendix Table A1 shows balance on stratification variables. Because recruiters' baseline CV scores are higher in the control group (Appendix Table A2), we control for CV grade in all specifications following Bruhn and McKenzie (2009); CVs were submitted prior to treatment and therefore cannot be affected by ChatGPT access.

To estimate the causal effect of ChatGPT access on cover letter evaluations and interview recommendations, we estimate:

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

where $ChatGPT_i$ indicates assignment to the ChatGPT-access condition and X_i includes stratification variables and any imbalanced baseline covariates. We include firm-by-recruiter fixed effects ($\mu_{f,r}$) and school fixed effects (μ_s). Standard errors are clustered at the job-seeker level (Abadie et al., 2022).[6] Estimation is by OLS unless the outcome is *Likelihood of Interview* (ordered logit) or *High Chance of Interview* (logit).

3 Job-seeker Experiment: Results

This section reports both pre-registered and exploratory results. The pre-registered outcomes are recruiter evaluations, browser history, and ChatGPT prompts (Table 1, Appendix Table A11, and Figure 2). All remaining analyses are exploratory.

3.1 Main Treatment Effects

Table 1 reports intention-to-treat estimates of ChatGPT access on recruiter-assessed cover letter quality and interview recommendations. Outcomes in columns (1)–(6) are standardized, so coefficients are interpreted in standard deviations. Access to ChatGPT increases the overall cover letter score by 0.222 standard deviations (column 1; $p < 0.10$). The improvement is concentrated in more standardized components—the introduction and closing—with effects of 0.253 and 0.281 standard

[6] Appendix Table A15 reports unclustered standard errors.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

deviations, respectively (columns 3 and 6; $p < 0.05$). We do not detect statistically significant effects on layout/clarity, experience, or motivation (columns 2, 4, and 5).

Despite higher average cover letter quality, we find no evidence that ChatGPT access increases interview recommendations. The estimated effect on the *Likelihood of Interview* measure is positive but statistically insignificant (column 7; $p = 0.380$), and the effect on the pre-registered indicator for a high interview likelihood (*Likelihood of Interview* > 3) is essentially zero (column 8; $p = 0.981$). Appendix Figure A2 shows that the implied marginal effects are small across the response categories, and Appendix Tables A7 and A8 confirm robustness to alternative inference and specification choices.[7]

To understand why improved cover letter scores do not increase interview recommendations, we examine which cover letter dimensions predict interview likelihood. Appendix Table A3 shows that the clarity and motivation components are the strongest predictors of interview recommendations. Consistent with Table 1, ChatGPT access does not measurably improve either component, providing a natural explanation for the null interview effect.

We next examine heterogeneity across the distribution of applicant quality. Using quantile regressions, we find that treatment effects are concentrated among lower-scoring applicants, consistent with evidence that LLMs disproportionately benefit weaker performers (Noy and Zhang, 2023; Dell'Acqua et al., 2023; Brynjolfsson, Li, and Raymond, 2025; Cowgill, Hernandez-Lagos, and Wright, 2025).[8] Figure 1 plots quantile-specific treatment effects for the total cover letter score (Panel A), introduction (Panel B), and closing (Panel C), showing declining effects across the distribution. In contrast, heterogeneous effects for layout, motivation, and experience are not statistically significant (Appendix Figure A1).

[7] We additionally verify that recruiters did not use LLMs to produce their evaluations by re-grading cover letters with ChatGPT under the same rubric (Appendix A.1.7). ChatGPT, blind to treatment status, assigns systematically higher scores to cover letters it helped generate, which is inconsistent with recruiter evaluations being produced by the same tool. This exercise also illustrates a potential evaluation bias if firms were to rely on LLM-based screening.

[8] We find no heterogeneous treatment effects by gender or by degree type; see Appendices A.1.3 and A.1.4. Furthermore, we document substantial variation of the correlation of within-firm recruiter evaluations, ranging from $\rho = 0.266 - 0.499$.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

3.2 ChatGPT's Impact on Writing Clarity

Table 1 shows that ChatGPT access improves overall cover letter evaluations but does not affect the recruiter-rated *Layout, Writing Quality, and Clarity* dimension. To interpret this null effect, we examine objective text measures that capture two distinct dimensions of writing quality: grammatical correctness and readability (Wiles, Munyikwa, and Horton, 2025).

Table 2 reports treatment effects on grammatical errors, readability, and length. ChatGPT access reduces the number of grammatical errors and the error rate by 37–39% (columns 1–2; $p < 0.01$), driven primarily by fewer spelling errors (Appendix Table A17; Appendix Figures A3 and A4). However, readability deteriorates: the Flesch Reading Ease score declines by 4.45 points (about 12%; column 3; $p < 0.05$), and the Flesch–Kincaid Grade Level increases modestly (column 4; $p = 0.11$). Consistent with reduced readability, ChatGPT increases average word length by 3% (column 6; $p < 0.01$) with no detectable effect on total word count (column 5). Appendix Figures A5 and A6 show the full distributions.

Taken together, these results indicate that ChatGPT improves surface-level technical correctness while making the text harder to read. This helps reconcile the null effect on recruiter-rated clarity: reduced errors and reduced readability operate in opposite directions. More broadly, the divergence between technical quality and readability suggests that generative AI can alter the nature of job-application signals in ways that are not necessarily decision-relevant for screening. This distinction is consistent with the idea that LLMs may raise the stylistic sophistication of application materials without increasing their informativeness about worker quality, an implication we quantify in the matching framework in Section 5.

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 the role of more effective prompting.

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 compre-

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

hensive 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 sub-sections), 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 A37 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 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 have notably different prompting 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, Munyikwa, and Horton, 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%

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

of users sought general strategy advice, 85% conducted company research, 60% made revision requests, and 17% asked for assistance 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. Appendix Table A11 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, Munyikwa, and Horton (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 A10).[9] Appendix Table A13 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.

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

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

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. Appendix Table A14 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.[10] 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 Appendix Table A14 suggests that the prompting guide did not amplify the effectiveness of the ChatGPT intervention, in line with recent studies (Bashardoust et al., 2024). Relatedly, Appendix Table A16 reports the correlational relationship between cover letter quality and number of ChatGPT prompts, with no persistent relationship found.

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,

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

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.

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 influence on recruiters' decisions.[11]

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.[12] The cover letters are drawn from the sample of cover letters that were written by job-seekers in the job-seeker experiment.[13] 84% of recruiters are employed either full- or part-time, 59% identify as female, with an average age of 37. 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. 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

[11] Among the recruiters, 65.3% are from the United Kingdom, while the remaining 34.7% come from other European countries. All recruiters work in Human Resources, are fluent in English, and have been active on Prolific within the past 90 days.

[12] Recruiters are paid a fixed fee. This is in line with Carvajal, Franco, and Isaksson (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).

[13] 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 ($p = 0.95$).

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

cover letters are written with the assistance of ChatGPT (akin to the setup of Cowgill, Hernandez-Lagos, and Wright, 2025). 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.

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 LLMs:

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

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 zero otherwise. $GPTwritten_r$ is an indicator variable equal to one 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 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.[14] 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 results presented in section 4.2 were pre-registered.

[14]Table A30 reports treatment effects without clustering.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

4.2 Results

Table 3 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 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 ($p = 0.5114$).[15] This provides evidence that while recruiters are aware that applicants use LLMs in their job applications, they are unable to identify which applicants actually use LLMs. 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 (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 eval-

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

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

uated 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.[16]

Taking advantage of our experimental design, we can gain further insights into what 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 similar variation in the quality of cover letters.

Appendix A.2.3 decomposes the average treatment effects from Table 3 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.

Table 4 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 for the cover letter being written with LLM assistance on the overall rating of the cover letters,

[16] 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.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

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 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 capture the effects of LLMs at a specific moment and under an exogenous adoption rate. To assess the resulting efficiency implications, we develop an assignment model with imperfect information. 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;

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Eeckhout and Kircher, 2018), we introduce LLMs into a static one-to-one assignment framework in which firms cannot observe workers' true abilities and instead rely on noisy signals.[17] Motivated by our first experiment, we model LLMs as a signal-enhancing technology that disproportionately improves the signals of lower-ability workers, generating a non-linear distortion in the distribution of signals observed by firms. Firms form Bayesian beliefs about workers' abilities and the likelihood that LLMs were used, and matches are determined through positive assortative matching (PAM) based on these beliefs. While LLMs may improve signals for some workers, the resulting compression and added uncertainty reduce the informativeness of signals and distort sorting, lowering aggregate matching efficiency relative to benchmarks with either perfect information or symmetric noise without LLMs.[18]

We focus on PAM because it closely reflects the labor markets relevant to our empirical setting—university graduates applying to entry-level, full-time positions. PAM implies that higher-ability workers match with higher-quality firms, concentrating talent in more selective jobs, while lower-ability workers sort into less demanding roles (Arcidiacono, Bayer, and Hizmo, 2010). This assumption is particularly plausible for the multinational firms in our first experiment, which operate in banking, insurance, and technology and compete for graduates from selective universities.

The heterogeneous effects documented in our experiments—where LLMs disproportionately inflate signals at the bottom of the distribution—interact directly with PAM. In such markets, signal distortion obscures the true ranking of candidates and undermines efficient sorting. Lower-ability

[17] Wiles, Munyikwa, and Horton (2025) also model hiring under imperfect information, but our frameworks differ along key dimensions. In Wiles, Munyikwa, and Horton (2025), workers are of two types (H and L), and only H-types benefit from algorithmic writing assistance through lower effort costs or the removal of "bad writing" shocks; as a result, AWAs increase signal dispersion. In contrast, consistent with our Experiment #1, LLMs in our model disproportionately benefit lower-ability workers, compressing the distribution of signals. Moreover, firms in Wiles, Munyikwa, and Horton (2025) are perfect substitutes and pay workers their expected productivity, whereas our model features heterogeneous firms and workers with positive assortative matching. This two-sided structure implies that LLM-induced signal compression generates mismatches and reduces aggregate output, while welfare effects in the partial-equilibrium model of Wiles, Munyikwa, and Horton (2025) are ambiguous, as gains for high-quality workers come at the expense of low-quality ones.

[18] Cowgill, Hernandez-Lagos, and Wright (2025) similarly model screening with noisy signals, but treat workers' ability and AI-induced signal gains as exogenously and jointly distributed, with an estimated slightly negative covariance implying larger gains for lower-type workers. Furthermore, AI usage is common knowledge to recruiters. In contrast, we microfound this relationship through firm–worker complementarity and allow AI use to be imperfectly observable, so that inference about AI adoption endogenously generates mismatch and welfare losses.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

applicants with inflated signals may displace better-suited candidates at high-quality firms, while firms, anticipating this distortion, may optimally discount application materials more broadly. The model allows us to quantify these effects by comparing equilibrium outcomes with and without LLM adoption. We calibrate the model using experimental estimates of signal distortion and moments from the literature on worker and firm heterogeneity (Teulings, 1995; Eeckhout and Kircher, 2018). The calibration indicates that welfare losses relative to the no-LLM benchmark can reach up to 0.8%, highlighting the scope for misallocation in high-stakes labor markets.

A closely related distortion arises in the education literature on grade inflation. Models such as Chan, Hao, and Suen (2007) show that grade inflation disproportionately benefits lower-ability students, weakening the informational content of grades and deteriorating aggregate allocation. In our setting, LLMs generate an analogous distortion by compressing labor market signals. Closest to our framework, Schwager (2012) embeds grade inflation into a one-to-one assignment model with imperfect information and shows that although low-ability students gain, aggregate output falls. We reach a similar conclusion: under imperfect information, the ability-dependent, nonlinear 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 $\mathcal{X} := [\underline{x}, \overline{x}]$. Bounded supports are assumed for simplicity and do not affect our results. 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:[19] $f(x', s') - f(x, s') \geq f(x', s) - f(x, s)\ \forall s' \geq s\ \forall x' \geq x$.

Assumption 1 implies that the marginal gain from matching with a higher-x firm is larger

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

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

for higher-s workers, and symmetrically that the marginal gain from a higher-s worker is larger for higher-x firms. This complementarity implies positive assortative matching: the best firms match with the best workers and the worst firms with the worst workers, as in standard production specifications in Schwager (2012) and Eeckhout and Kircher (2018).

Firms do not directly observe workers' true abilities. Instead, each worker sends a noisy signal of her ability to all firms. After observing these signals, firms form Bayesian estimates of worker quality, denoted by $\tilde{s}$, and match assortatively based on these estimates. The assignment is a function $x = \sigma(\tilde{s})$, which maps each estimated quality $\tilde{s}$ to a firm x. In equilibrium, firms and workers are matched one-to-one.[20] The assignment rule is therefore deterministic:

$$x = \sigma(\tilde{s}) = G_x^{-1}\big(G_{\tilde{s}}(\tilde{s})\big),$$

where $G_{\tilde{s}}$ denotes the distribution of estimated worker qualities. All randomness in matching thus comes from noise in the signal, which makes $\tilde{s}$ deviate from s. Conditional on a worker's realized signal, their match is deterministic, as in a setting where each worker sends a single application containing their signal to all firms.

We assume that signals take the following form: a fraction p of workers sends a noisy linear signal. These are workers who either do not have access to, or choose not to adopt, LLM technology. The noise has mean zero and is uncorrelated with true ability. The remaining fraction $1-p$ adopts LLMs, which improve the quality of their signals.[21] For a worker of true quality s, the signal y observed by firms is

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

where e is a mean-zero random variable. We impose the following assumption on h, motivated by our first experiment, where LLMs primarily improve signals for low-ability workers while leaving

[20] 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.

[21] 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.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

high-ability workers essentially unaffected:

Assumption 2. $y_2 = h(y_1)$ is a continuous function that satisfies (i) $h(y_1) \geq y_1$ for all y_1, (ii) $0 < h'(y_1) < 1$ for all y_1, and (iii) $\overline{y} := h(\overline{s} + \overline{e}) = \overline{s} + \overline{e}$.

Assumption 2 implies that the signal improvement is larger for low y_1 than for high y_1, as in our job-seeker experiment (see Figure A15 for an illustration). A key feature of the information structure is that firms observe only y and do not know whether it was generated with LLM assistance. This aligns with our second experiment, where recruiters cannot reliably identify LLM usage.

Our transformation function h echoes the "grade-inflation" mapping in Schwager (2012): the incremental gain in the observed signal is largest for applicants with the lowest baseline ability. This contrasts with the clarity view in Wiles, Munyikwa, and Horton (2025), in whose experiment, algorithmic writing assistants act as noise reducers, improving writing quality for every applicant and sharpening the signal without substantially altering the relative ranking of applicants. Our experimental evidence points to a different mechanism. Because the quality boost diminishes with baseline ability, h compresses the signal distribution: low-ability applicants improve substantially while high-ability applicants benefit only marginally. When LLM usage is imperfectly observable, this creates additional uncertainty about how observed text maps into underlying ability. Capturing this non-linear distortion is therefore essential for assessing its consequences for sorting and welfare.

From the signal y, firms form the Bayesian estimate $\tilde{s}(y) = \mathbb{E}[s|Y = y]$ without ex ante knowing whether the signal was generated by LLMs. Firms can nonetheless infer the probability of LLM usage from the observed signal. By the law of total expectations,

$$\tilde{s}(y) = \omega(y)\mathbb{E}[s|Y_1 = y] + (1 - \omega(y))\mathbb{E}[s|Y_2 = y], \tag{3}$$

where $\omega(y)$ is the posterior probability of LLM usage conditional on y. By Bayes' rule,

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

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.

If h satisfies Assumption 2, then Y_2 takes values above $\underline{y}_2 := h(\underline{s}+\underline{e})$. For signals $y < \underline{y}_2$, $g_{y_2}(y) = 0$, so $\omega(y) = 1$ and $\tilde{s}(y) = \mathbb{E}[s|Y_1 = y]$. That is, sufficiently low signals are known to come from non-users. For $y \geq \underline{y}_2$, firms use (3) to form beliefs over worker quality.

Let $\hat{s}(y) := \mathbb{E}[s|Y_1 = y]$ denote the posterior mean in an economy without LLMs, where all firms know signals come from Y_1. Since h is monotone and invertible, $\mathbb{E}[s|Y_2 = y] = \mathbb{E}[s|Y_1 = h^{-1}(y)] = \hat{s}(h^{-1}(y))$.[22] Therefore,

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

Since $h^{-1}(y) \leq y$, we have $\tilde{s}(y) < \hat{s}(y)$ for $\underline{y}_2 \leq y < \overline{y}$ (Figure 3). Thus, over the range where signals may be LLM-augmented, firms discount observed text relative to the no-LLM benchmark, making signals less informative.

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

$$V_L = \int_{\mathcal{S}} \int_{\mathcal{S}} f(G_x^{-1}(G_{\tilde{s}}(\tilde{s})), s) \mathrm{d}G_{\tilde{s}|s}(\tilde{s}|s) \mathrm{d}G_s(s). \tag{4}$$

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 aggregate value in the economy with imperfect information and LLMs (V_L) is at most equal to the value in the imperfect-information economy without LLMs.[23] The intuition is that while LLMs can improve some signals, imperfect observability leads firms to discount signals over the range where augmentation is possible. This induces mismatches, and with concavity and complementarity in f, lowers aggregate output relative to the no-LLM benchmark.

Calibration. To quantify the efficiency losses from LLMs under imperfect information, we calibrate the model and simulate output losses. We assume a CES production function for the value

[22] See Appendix E.3 for details.
[23] See Appendix E for details and proof.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

of a firm–worker match, following Eeckhout and Kircher (2018), $f(x, s) = [\alpha x^{\rho} + (1 - \alpha)s^{\rho}]^{1/\rho}$. Table 5 reports the calibrated parameters and target moments.

Figure 4 plots efficiency losses induced by LLMs as a percentage relative to an economy with positive assortative matching and imperfect information but *without* LLMs (denoted by V_I). The horizontal axis reports p, the share of job-seekers who do not use LLMs. Efficiency losses peak at around 0.8% when approximately 80% of applicants use LLMs. In our job-seeker and recruiter experiments, where LLM usage rates are 47% and 50%, respectively, the implied efficiency loss is approximately 0.4%. A key feature of Figure 4 is that efficiency losses are non-monotonic in LLM adoption. When LLM usage is moderate, imperfect observability generates substantial uncertainty and misallocation. As adoption becomes nearly universal, however, this uncertainty diminishes: firms internalize that most applicants use LLMs and adjust their inference accordingly. In the limit where all applicants use LLMs ($p = 1$), firms perfectly invert the signal transformation $h(\cdot)$, so that $\tilde{s}(y) = \hat{s}(y)$ for all y, and efficiency losses disappear.

The figure also shows that LLM-induced misallocation is asymmetric around $p = 0.5$. This asymmetry reflects the distortionary effect of LLMs on signals: LLMs shift the left tail of the signal distribution upward while leaving the right tail largely unchanged, compressing signals and reducing their informativeness. As the share of LLM users increases, this compression intensifies and output falls, until adoption becomes sufficiently widespread that firms can again infer worker quality more accurately.

In Appendix E, we extend the model to allow LLM adoption to be correlated with underlying ability s. This may arise if higher-ability applicants have better access to, or are more willing to use, LLMs, or if lower-ability applicants anticipate larger gains from adoption. We model this by allowing p to vary linearly with s, holding the average adoption rate fixed at 0.5. The results show that aggregate output increases as LLM adoption becomes more positively correlated with ability, consistent with the model's mechanism: when higher-ability applicants are more likely to adopt, fewer low-ability signals are distorted, reducing mismatch and improving allocation.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.

6 Conclusion

The emergence of Large Language Models has the potential to significantly impact productivity and reshape labor market dynamics. LLMs are already widely adopted, with Bick, Blandin, and Deming (2025) documenting that nearly 40% of the US working-age population uses them, and 23% of employees using them at least once per week. However, the influence of LLM usage by job-seekers on labor market signals and subsequent matching efficiency remains largely unexplored. This paper addresses this gap through two field experiments involving job-seekers and recruiters, whose findings inform a calibrated 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. These gains, however, do not translate into higher interview probabilities, as the improvements are concentrated in less critical and less personalized sections of the cover letter. Moreover, while ChatGPT reduces spelling and grammar errors, it simultaneously lowers readability by increasing linguistic complexity. Our second experiment shows that recruiters are no better than chance at detecting LLM usage, and that disclosure interacts with recruiters' expectations: high-quality cover letters written without LLM assistance are evaluated more positively when recruiters are aware of this, indicating a premium placed on clearly human-written applications.

Taken together, our experimental results reveal important non-linearities in how LLMs affect job-seekers. Lower-ability applicants benefit disproportionately from LLM assistance, while applicants already producing high-quality cover letters experience little improvement and may be disadvantaged when recruiters become cautious about AI-generated content. These patterns motivate our model, which embeds LLM-induced signal compression into an imperfect-information matching framework with firm-worker complementarity. The calibration shows that widespread LLM adoption can reduce aggregate welfare by distorting matching: lower-ability applicants are matched to overly demanding positions, while higher-ability applicants are displaced into less suitable roles. Although some firms or workers may benefit from this misallocation, the net effect is a decline in aggregate welfare relative to a benchmark with imperfect information but no LLM usage.

These findings are particularly relevant given the early stage of generative AI adoption, limited

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.

regulation, and uncertain future norms. If policies mandating explicit disclosure of LLM use—such as those currently discussed in the European Union—become widespread, the conditions in our *Full Information* treatment may approximate future hiring environments. More broadly, our results extend to settings beyond labor markets in which evaluators rely on written materials to assess applicants' ability, such as university admissions.[24] While domains differ, the core features—written signals, imperfect observability of LLM use, and evaluator inference—are similar.

Finally, because our study relies on the freely accessible ChatGPT 3.5, our estimates likely represent a conservative lower bound on the effects of generative AI. As LLM capabilities continue to improve, understanding how these tools reshape signals, expectations, and matching remains an important avenue for future research.

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

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Data Availability

The data and code replicating all tables and figures in this article, together with a README file documenting their contents and use, can be found in Abbas Nejad et al. (2026) in the Harvard Dataverse, https://doi.org/10.7910/DVN/LPIREO.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2022. "When Should You Adjust Standard Errors for Clustering?" *The Quarterly Journal of Economics* 138 (1):1–35.

Abbas Nejad, Kian, Giuseppe Musillo, Till Wicker, and Niccolò Zaccaria. 2026. "Replication Package for: 'Labor Market Signals: The Role of Large Language Models'." Harvard Dataverse, https://doi.org/10.7910/DVN/LPIREO.

Ameriks, John, Joseph Briggs, Andrew Caplin, Matthew D. Shapiro, and Christopher Tonetti. 2020. "Long-term-care Utility and Late-in-Life Saving." *Journal of Political Economy* 128 (6):2375–2451.

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart. 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." URL https://persportaal.anp.nl/artikel/544fd858-dc4e-46b7-ac69-e91cafffc4f8/kwart-organisaties-screent-sociale-media-van-sollicitant. Accessed: 2024-11-25.

———. 2024. "Werkgevers gedogen sollicitatiebrief en cv van ChatGPT." URL https://persportaal.anp.nl/artikel/d694a17f-4e6a-454d-b6f5-1498bf1541a8/werkgevers-gedogen-sollicitatiebrief-en-cv-van-chatgpt. Accessed: 2024-11-25.

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

Arellano-Bover, Jaime. 2024. "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size." *Journal of Labor Economics* 42 (2):549–589.

Autor, David H. 2001. "Wiring the Labor Market." *Journal of Economic Perspectives* 15 (1):25–40. URL https://www.aeaweb.org/articles?id=10.1257/jep.15.1.25.

Avery, Mallory, Andreas Leibbrandt, and Joseph Vecci. 2024. "Does artificial intelligence help or hurt gender diversity? Evidence from two field experiments on recruitment in tech." CESifo Working Paper.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Awuah, Kobbina, Urša Krenk, and David Yanagizawa-Drott. 2025. "Automation with generative AI? Evidence from a teacher hiring pipeline." Working Paper.

Bashardoust, Amirsiavosh, Yuanjun Feng, Dominique Geissler, Stefan Feuerriegel, and Yash Raj Shrestha. 2024. "The Effect of Education in Prompt Engineering: Evidence from Journalists." URL https://arxiv.org/abs/2409.12320.

Bassi, Vittorio and Aisha Nansamba. 2021. "Screening and Signalling Non-Cognitive Skills: Experimental Evidence from Uganda." *The Economic Journal* 132 (642):471–511.

Bick, Alexander, Adam Blandin, and David J. Deming. 2025. "The Rapid Adoption of Generative AI." URL https://static1.squarespace.com/static/60832ecef615231cedd30911/t/679bb3c5f3307140dd839d94/1738257349610/BBD_GenAI_Jan2025_main.pdf. Working Paper.

Böhm, Robert, Martin Jörling, Lisa Reiter et al. 2023. "People devalue generative AI's competence but not its advice in addressing societal and personal challenges." *Communications Psychology* 1:32. URL https://doi.org/10.1038/s44271-023-00032-x.

Bohren, Noah, Rustamdjan Hakimov, and Rafael Lalive. 2024. "Creative and Strategic Capabilities of Generative AI: Evidence from Large-Scale Experiments." Tech. Rep. DP No. 17302, IZA Institute of Labor Economics, Bonn, Germany. URL https://docs.iza.org/dp17302.pdf. IZA Discussion Paper Series; ISSN: 2365-9793.

Brañas-Garza, Pablo, Lorenzo Estepa-Mohedano, Diego Jorrat, Victor Orozco, and Ericka Rascón-Ramírez. 2021. "To pay or not to pay: Measuring risk preferences in lab and field." *Judgment and Decision Making* 16 (5):1290–1313.

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

Bruhn, Miriam and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics* 1 (4):200–232.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at Work." *The Quarterly Journal of Economics* URL https://doi.org/10.1093/qje/qjae044.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Caplin, Andrew, David J Deming, Shangwen Li, Daniel J Martin, Philip Marx, Ben Weidmann, and Kadachi Jiada Ye. 2024. "The ABC's of Who Benefits from Working with AI: Ability, Beliefs, and Calibration." Tech. rep., National Bureau of Economic Research.

Carranza, Eliana, Robert Garlick, Kate Orkin, and Neil Rankin. 2022. "Job Search and Hiring with Limited Information about Workseekers' Skills." *American Economic Review* 112 (11):3547–83. URL https://www.aeaweb.org/articles?id=10.1257/aer.20200961.

Carvajal, Daniel, Catalina Franco, and Siri Isaksson. 2024. "Will Artificial Intelligence Get in the Way of Achieving Gender Equality?" Discussion Paper 03, NHH Dept. of Economics. URL https://ssrn.com/abstract=4759218.

Chan, William, Li Hao, and Wing Suen. 2007. "A sitnaling theory of grade inflation." *International Economic Review* 48 (3):1065–1090. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2354.2007.00454.x.

Chaturvedi, Sugat and Rochana Chaturvedi. 2025. "Who gets the callback? Generative AI and gender bias." *arXiv preprint arXiv:2504.21400* .

Choi, Jonathan H. and Daniel Schwarcz. 2024. "AI Assistance in Legal Analysis: An Empirical Study." *Journal of Legal Education* .

Cowgill, Bo, Pablo Hernandez-Lagos, and Nataliya Wright. 2025. "Does AI Cheapen Talk? Theory and Evidence From Global Entrepreneurship and Hiring." Columbia Business School Research Paper No. 4702114, Columbia Business School Research Paper Series.

Criddle, Cristina and Delphine Strauss. 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].

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani. 2023. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality."

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Dell'Acqua, Fabrizio. 2022. "Falling Asleep at the Wheel: Human/AI Collaboration in a Field Experiment on HR Recruiters." Working paper. Unpublished manuscript.

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

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market*." *The Quarterly Journal of Economics* 132 (4):1593–1640.

Deming, David J, Christopher Ong, and Lawrence H Summers. 2025. "Technological Disruption in the Labor Market." Working Paper 33323, National Bureau of Economic Research. URL http://www.nber.org/papers/w33323.

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

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. "Gpts are gpts: An early look at the labor market impact potential of large language models." *arXiv preprint arXiv:2303.10130* .

Englmaier, Florian, Stefan Grimm, Dominik Grothe, David Schindler, and Simeon Schudy. 2024. "The effect of incentives in nonroutine analytical team tasks." *Journal of Political Economy* 132 (8):2695–2747.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2023. "The preference survey module: A validated instrument for measuring risk,time, and social preferences." *Management Science* 69 (4):1935–1950.

Hoffman, Mitchell, Lisa B Kahn, and Danielle Li. 2017. "Discretion in Hiring." *The Quarterly Journal of Economics* 133 (2):765–800.

Kadoma, Kowe, Danae Metaxa, and Mor Naaman. 2024. "Generative AI and Perceptual Harms: Who's Suspected of using LLMs?" URL https://arxiv.org/abs/2410.00906.

Kurlat, Pablo and Florian Scheuer. 2020. "Signalling to Experts." *The Review of Economic Studies* 88 (2):800–850.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies* 76 (3):1071–1102.

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

Otis, Nicholas, Rowan Clarke, Solène Delecourt, David Holtz, and Rembrand Koning. 2024. "The Uneven Impact of Generative AI on Entrepreneurial Performance." *Management Science* .

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot." URL https://arxiv.org/abs/2302.06590.

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, Robert. 2012. "Grade inflation, social background, and labour market matching." *Journal of Economic Behavior & Organization* 82 (1):56–66. URL https://www.sciencedirect.com/science/article/pii/S0167268111003131.

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

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

Stango, Victor and Jonathan Zinman. 2023. "We Are All Behavioural, More, or Less: A Taxonomy of Consumer Decision-Making." *The Review of Economic Studies* 90 (3):1470–1498.

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

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

The Guardian. 2023. "ChatGPT and AI: How they can help disadvantaged college applicants in the era of affirmative action." URL https://www.theguardian.com/education/2023/aug/27/chatgpt-ai-disadvantaged-college-applicants-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." URL https://www.bls.gov/news.release/archives/work_12192023.htm. Accessed: 2024-12-08.

Wiles, Emma, Zanele Munyikwa, and John Horton. 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, Yunhao and Renée Gosline. 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.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.

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.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Table 2: 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.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Table 3: Main Results: 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.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Table 4: 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.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Table 5: Calibration Parameters

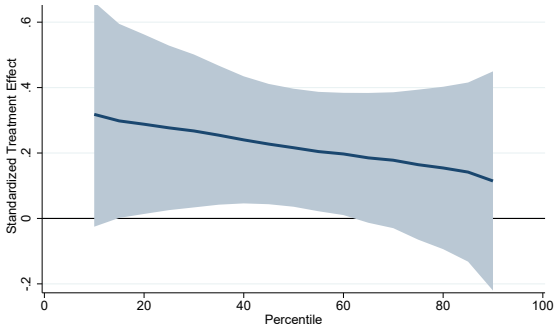
Parameter	Source	Value
Production Function	Eeckhout and Kircher (2018)	$\alpha = 0.5; \rho = 0.5$
Distribution of Worker Type	Teulings (1995)	$N(5.59, 1.29)$
Distribution of Firms Type	Teulings (1995)	truncated $N(5.59, 1.29)$
Slope of h function	Experiment #1	0.61

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.

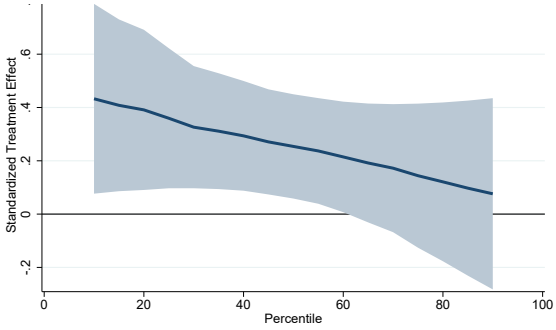
This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Figure 1. Quantile Regressions

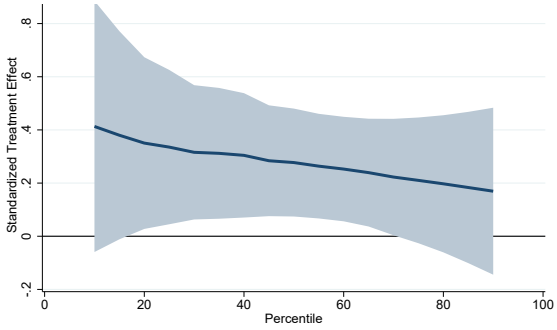
(A) Cover Letter: Total Score



(B) Cover Letter: Introduction



(C) Cover Letter: Closing

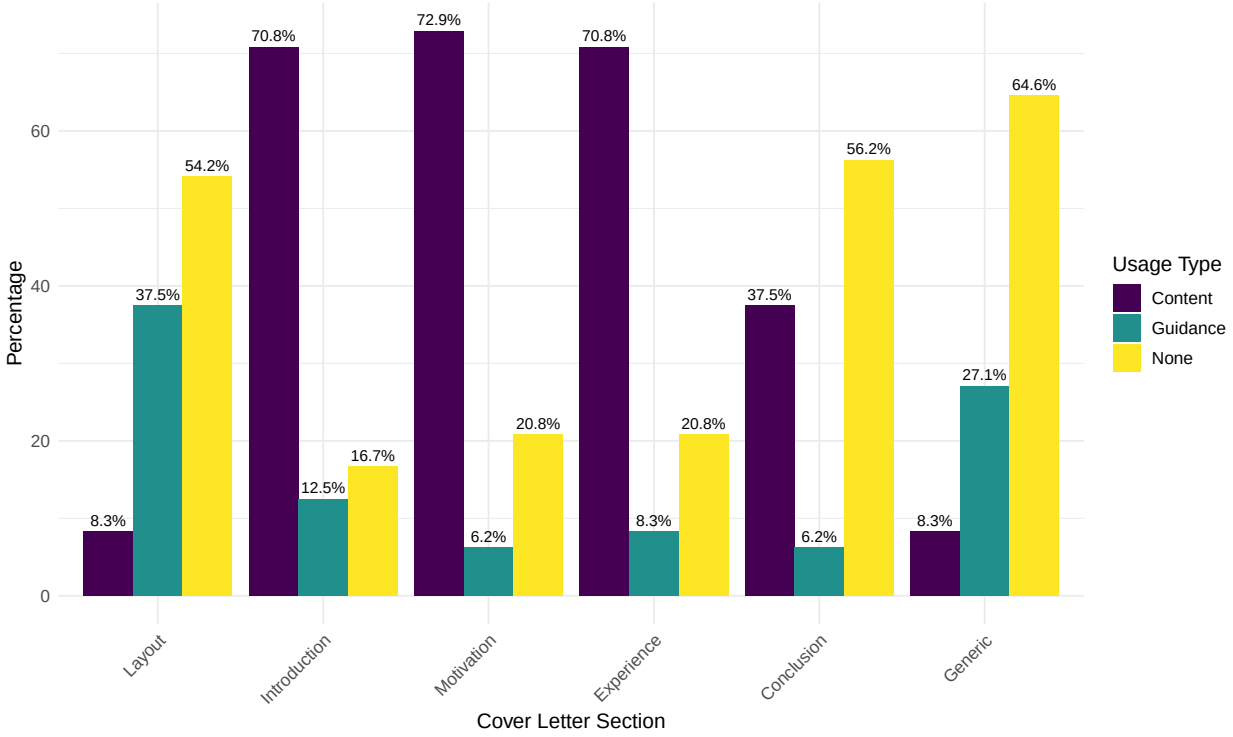


Notes: Each panel plots the estimated treatment effect of ChatGPT on cover letters across quantiles of the outcome distribution, obtained from quantile regressions. The outcomes correspond to the three main sub-components of the cover letter—Total Score, Introduction, and Closing—as described in Appendix 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.

Alt text: Three-panel figure showing quantile regression estimates of the effect of LLM assistance on cover letter quality. Panel (A) plots the standardized treatment effect on the total cover letter score across outcome percentiles, showing positive effects at lower percentiles that steadily decline toward zero at higher percentiles. Panel (B) shows a similar declining pattern for the introduction section, with larger positive effects at the bottom of the distribution and smaller effects at the top. Panel (C) plots the effect on the closing section, displaying smaller and flatter positive effects that decrease modestly across percentiles. Shaded areas in all panels indicate 95% confidence intervals.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication. The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.

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

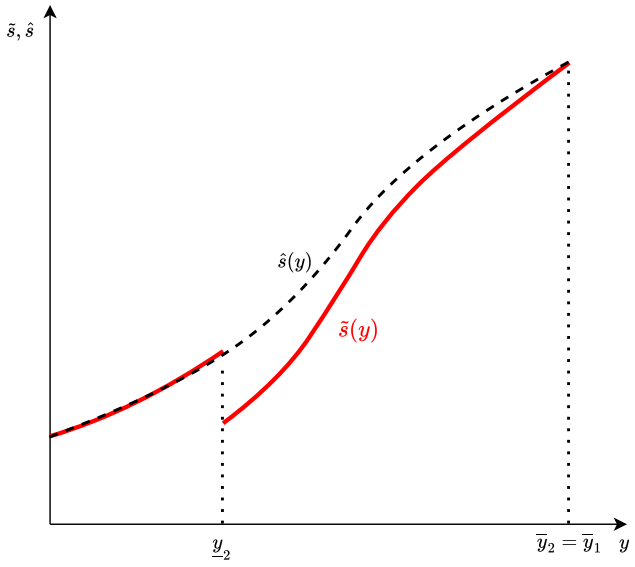


Notes: The figure shows the distribution of ChatGPT usage across cover letter sections. Each bar reflects the share of conversations where the user requested Content (i.e., text generation or improvement), Guidance (i.e., structural or formatting advice), or showed No engagement with the section (None).

Alt text: Grouped bar chart showing how ChatGPT conversations are distributed across cover letter sections. For each section—layout, introduction, motivation, experience, conclusion, and generic requests—the bar displays the percentage of interactions classified as content generation or editing, structural guidance, or no engagement. Most interactions involve content generation, guidance requests are relatively rare, and some sections show substantial shares with no engagement. Introduction and conclusion sections exhibit the highest concentration of content-related requests.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.

Figure 3. Comparison of $\tilde{s}$ and $\hat{s}$

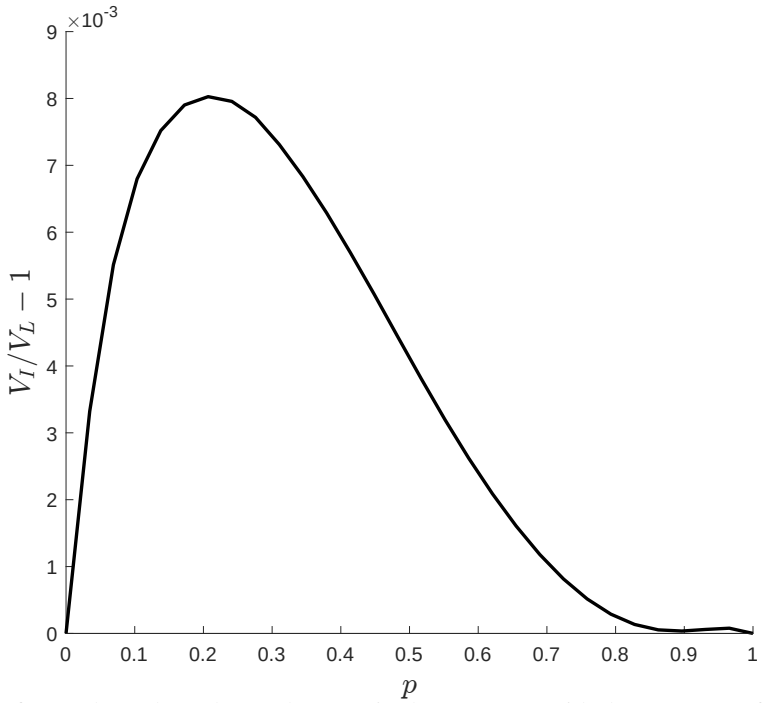


Notes: The figure plots firms' expectations of worker ability without LLMs, $\hat{s}(y)$, and when LLMs are available, $\tilde{s}(y)$. Since LLMs improve signal quality, firms know that low signals must come from non-users, so the two expectations coincide for small y.

Alt text: Line diagram comparing firms' expected worker ability as a function of the observed signal. The dashed black curve shows expectations without large language models, denoted $\hat{s}(y)$, and the solid red curve shows expectations when large language models are available, denoted $\tilde{s}(y)$. The two curves overlap for low signal values, indicating identical beliefs, and diverge for higher signals, with the LLM curve lying above, reflecting improved signal quality and higher inferred ability.

This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.

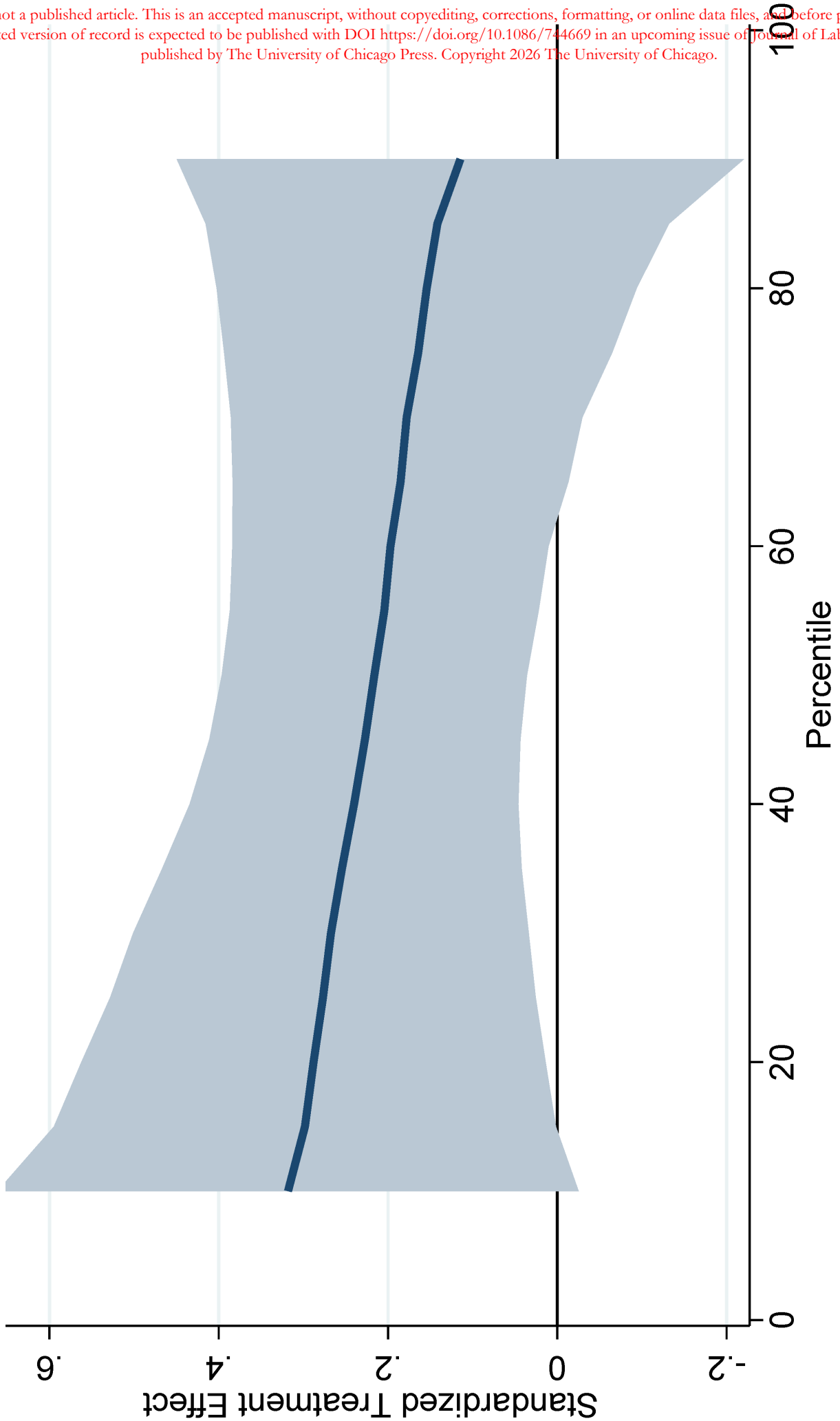
Figure 4. Sensitivity of Misallocation to p



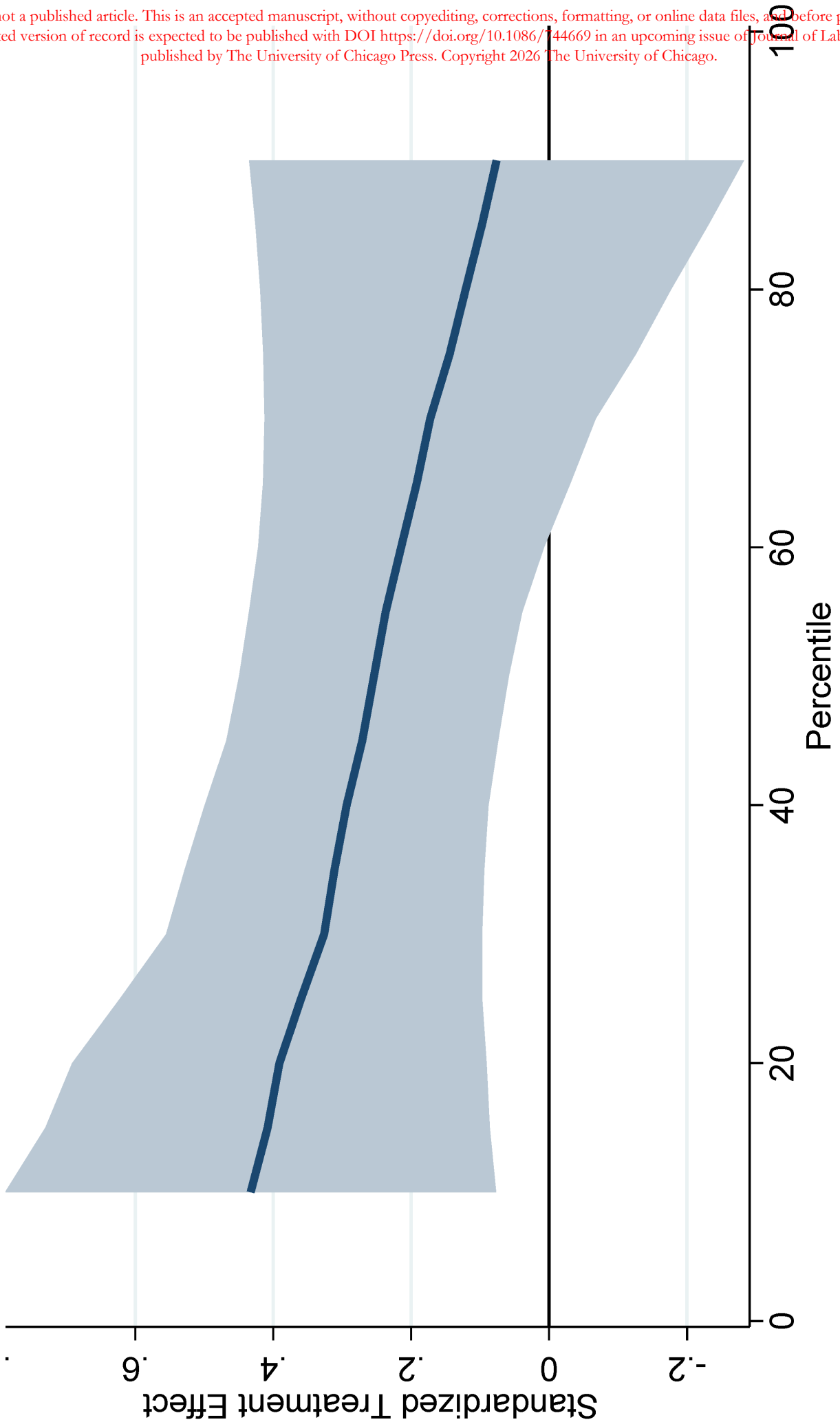
Notes: The figure shows how the total output in the economy with the presence of LLM technology V_L (calculated using equation (4)) differs from the economy where there are imperfect labor market signals but no LLM technology V_I (see Appendix E for details) when the proportion p of applicants who do *not* use LLM technology varies over $[0, 1]$.

Alt text: Line plot showing how the difference in total output between an economy with LLM technology and one without varies with p, the share of applicants who do not use LLMs (horizontal axis from 0 to 1). The vertical axis reports the output difference. The relationship is inverse-U shaped: the output difference is near zero when p is close to 0, rises to a maximum at an intermediate value of p (around $p \approx 0.2$), and then declines toward zero again as p approaches 1, indicating the largest output losses occur at intermediate non-adoption rates.

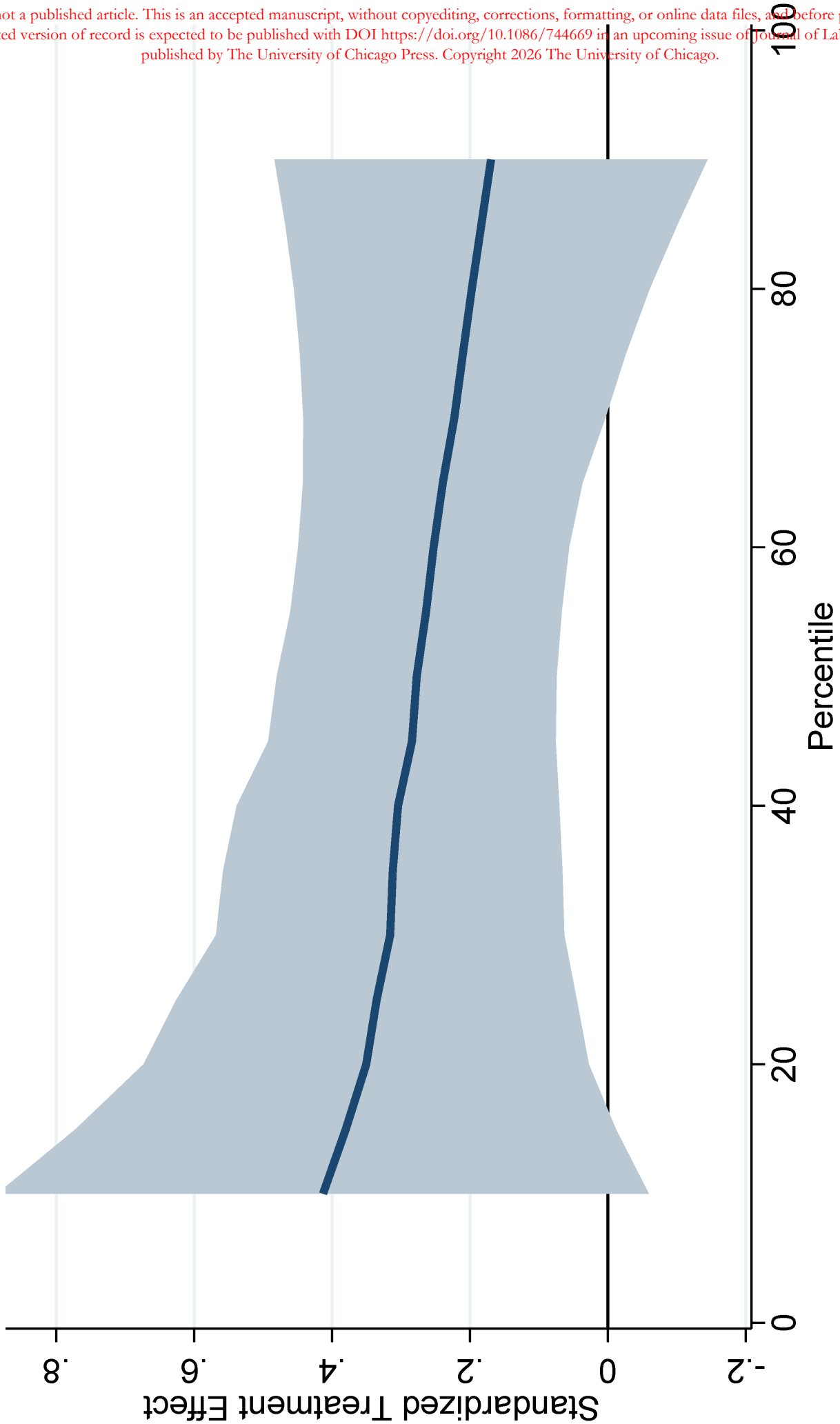
This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.



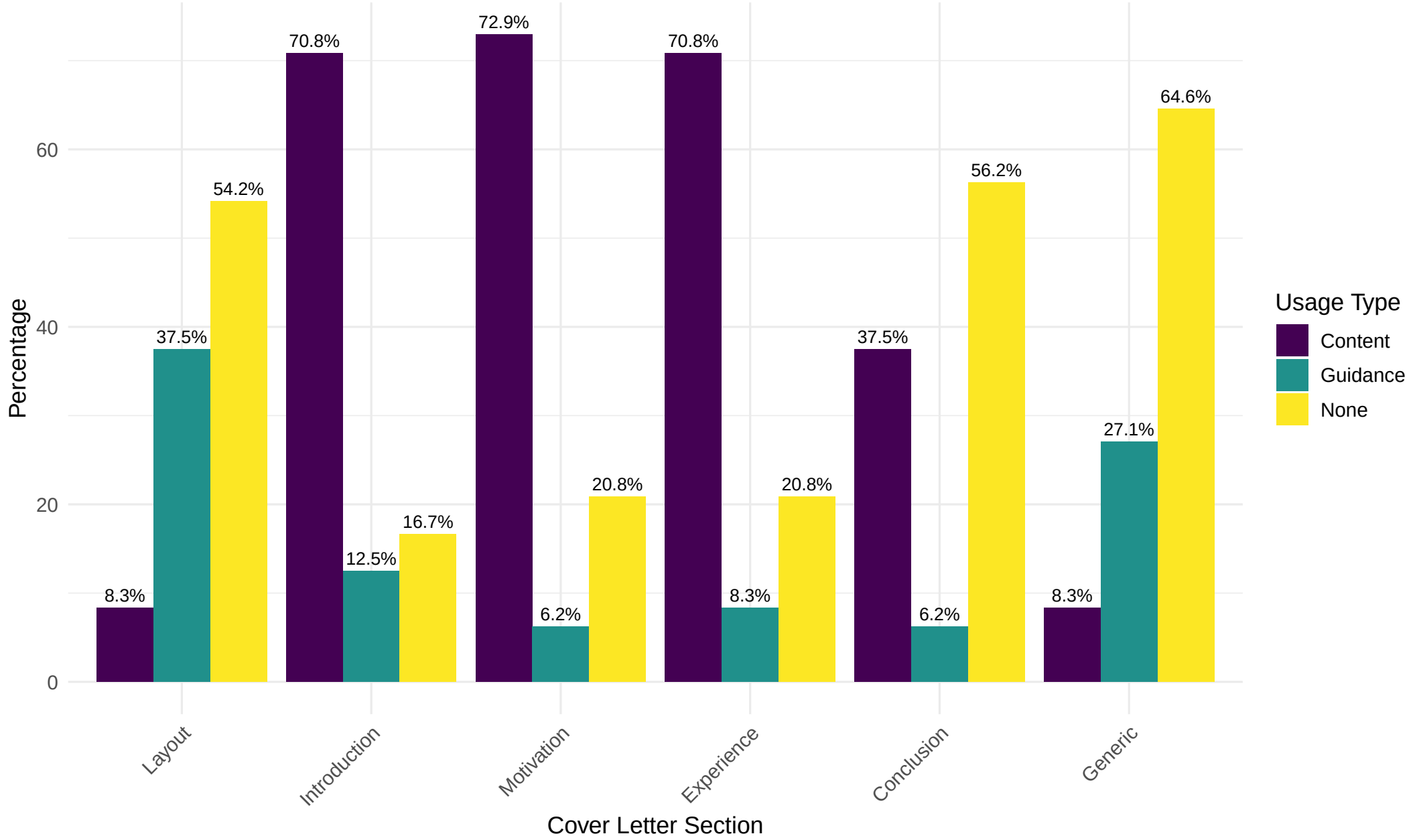
This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.



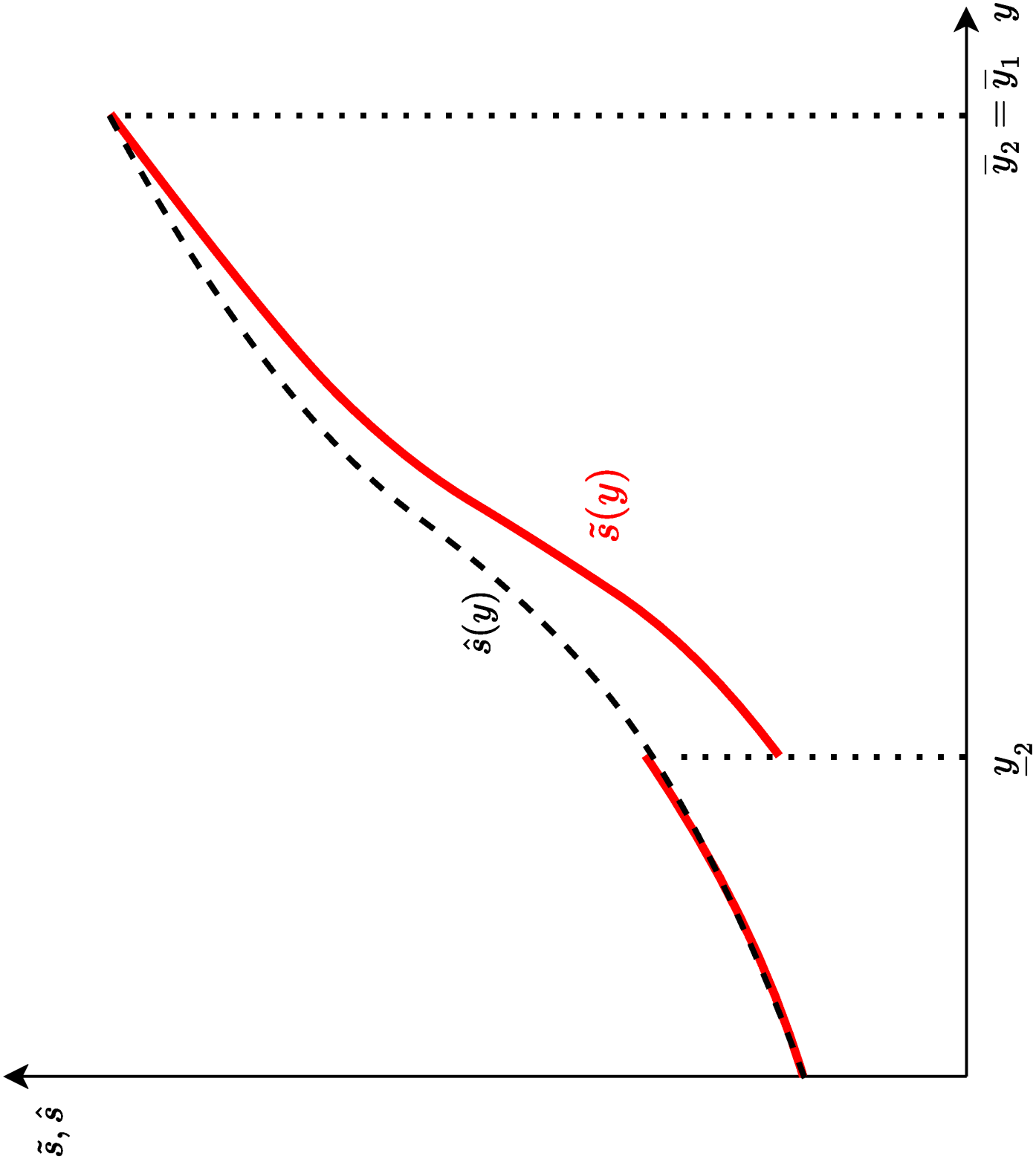
This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, used before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.



This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics, published by The University of Chicago Press. Copyright 2026 The University of Chicago.



This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.



This is not a published article. This is an accepted manuscript, without copyediting, corrections, formatting, or online data files, and before publication.
The completed version of record is expected to be published with DOI https://doi.org/10.1086/744669 in an upcoming issue of Journal of Labor Economics,
published by The University of Chicago Press. Copyright 2026 The University of Chicago.

