Labor Market Signals: The Role of Large Language Models

Kian Abbas Nejad Giuseppe Musillo Till Wicker Niccolò Zaccaria

Tilburg University

Motivation: LLMs and labor market signals

- **How do LLMs affect hiring signals?**
- In many hiring settings, firms cannot directly observe worker productivity or fit.
- Instead, they rely on **signals** such as CVs and cover letters.
- LLMs may improve these signals by raising writing quality.
- But they may also make texts **less informative** about underlying ability.

Motivation: LLMs and labor market signals

- **How do LLMs affect hiring signals?**
- In many hiring settings, firms cannot directly observe worker productivity or fit.
- Instead, they rely on **signals** such as CVs and cover letters.
- LLMs may improve these signals by raising writing quality.
- But they may also make texts **less informative** about underlying ability.
- This creates a central question: **do better-looking signals improve matching, or distort it?**

Why cover letters?

- Cover letters are still common in entry-level hiring.
- They are meant to reveal:
  - motivation
  - communication ability
  - fit with the job
  - fit with the organization
- This is exactly the kind of persuasive writing task where LLM effects are ambiguous ex ante.

Conceptual tension

- **Clarity view:** LLMs reduce information frictions.
- **Signaling view:** LLMs weaken the link between text quality and applicant ability.

- The paper asks which force dominates in actual recruiter evaluations.

What can LLMs do to a labor market signal?

Quality channel

- Better grammar
- More polished writing
- Stronger structure
- Bigger gains for weaker applicants

Prediction: better cover letters

$=$

Signal channel

- Less personalization
- Compression of quality differences
- Harder to infer true applicant quality

Prediction: weaker screening value

Our question: Do LLMs help applicants get interviews, or do they mainly inflate signals?

This paper

1. Do LLMs improve the quality of job-seekers' cover letters?
2. Do those improvements translate into interview recommendations?
3. How do recruiters respond when LLM use is disclosed?
4. What does this imply for **signal informativeness** and matching?

Main findings

- Access to ChatGPT improves cover letter quality by **0.22 standard deviations**.
- Gains are concentrated among **lower-quality applicants**.
- Improvements appear in **standardized sections** such as the introduction and closing.
- ChatGPT does **not** improve the dimensions recruiters care about most: **motivation and clarity**.
- Therefore ChatGPT access does **not** increase interview recommendations.

Main findings

- Access to ChatGPT improves cover letter quality by **0.22 standard deviations**.
- Gains are concentrated among **lower-quality applicants**.
- Improvements appear in **standardized sections** such as the introduction and closing.
- ChatGPT does **not** improve the dimensions recruiters care about most: **motivation and clarity**.
- Therefore ChatGPT access does **not** increase interview recommendations.
- When recruiters are explicitly informed about LLM usage, they place greater value on **high-quality human-written** cover letters.
- A calibrated assignment model implies **welfare losses through mismatch**.

Contribution

- **Labor market signaling**
  ⇒ Spence (1973), Kurlat and Scheuer (2020), Bassi and Nansamba (2021), Carranza et al. (2022)
  Contribution: shows that LLMs can **inflate and compress** written signals.
- **Technology and matching**
  ⇒ Autor (2001), Wiles et al. (2025), Cowgill et al. (2024)
  Contribution: distinguishes **generative AI** from editing tools by showing that LLMs alter the **content and informativeness** of the signal.
- **Generative AI and productivity**
  ⇒ Noy and Zhang (2023), Dell'Acqua et al. (2023), Caplin et al. (2024)
  Contribution: in a persuasive writing task, LLM gains are concentrated where writing is **generic rather than individualized**.

Experiment I: design

- Field experiment with **137 students** from Tilburg University and Utrecht University.
- Four employers: **Philips, PwC, Rabobank, VodafoneZiggo**.
- Students apply to a real entry-level vacancy and have **one hour** to write a cover letter.
- Applications are pseudonymized and evaluated by recruiters from the relevant firm.

Experiment I: design

- Field experiment with **137 students** from Tilburg University and Utrecht University.
- Four employers: **Philips, PwC, Rabobank, VodafoneZiggo**.
- Students apply to a real entry-level vacancy and have **one hour** to write a cover letter.
- Applications are pseudonymized and evaluated by recruiters from the relevant firm.
- Treatment:
  - **Control:** no access to LLM websites
  - **ChatGPT:** access to GPT-3.5 through a researcher-provided account
- Recruiters are **blind to treatment status**.

What is observed?

- Recruiter evaluations of:
  - total cover letter quality
  - layout / clarity
  - introduction
  - experience
  - motivation
  - closing
  - interview likelihood

- Browser histories
- Full ChatGPT conversation logs
- This lets us observe both:
  - the **quality of the signal**
  - the **process through which it is produced**

Empirical strategy

Main specification

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

- $ChatGPT_i$: assigned access to ChatGPT
- Controls include stratification variables and imbalanced baseline covariates
- Firm-by-recruiter fixed effects and school fixed effects
- Standard errors clustered at the job-seeker level

Interpretation: causal effect of **LLM access** on cover letter quality and interview recommendations.

Main results: ChatGPT improves the letter, but not interview chances

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Cover Letter					
	Total	Layout	Intro	Experience	Motivation	Closing	Likelihood of Interview	High Chance of 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 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

- ChatGPT improves **overall quality**, especially the **introduction** and **closing**.
- But there is **no detectable increase** in interview recommendations.

Why no interview effect?

Recruiters care most about:

- **Motivation**
- **Clarity / layout / writing quality**

But ChatGPT does not improve either one enough.

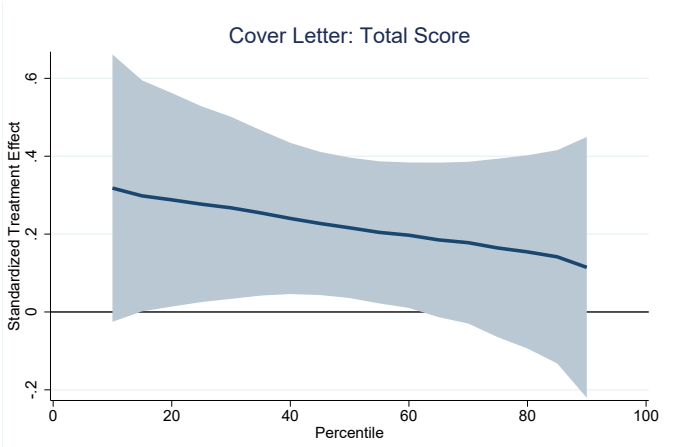
- Motivation remains hard to personalize.
- Clarity is unchanged on net.

Predictor	Interview likelihood
CL: Layout	0.525^{**}
CL: Motivation	0.823^{***}

So LLMs improve the parts that are easier to standardize, not the parts that drive interview calls.

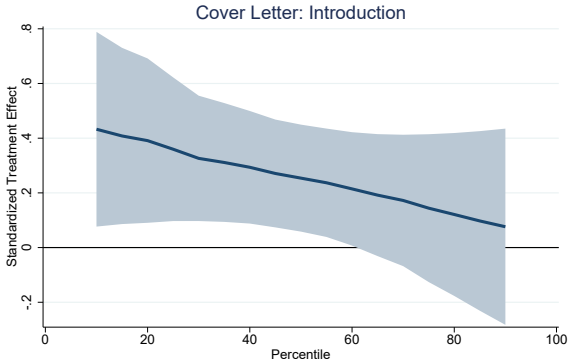
Determinants of interview likelihood

Heterogeneity: who benefits from LLM access?

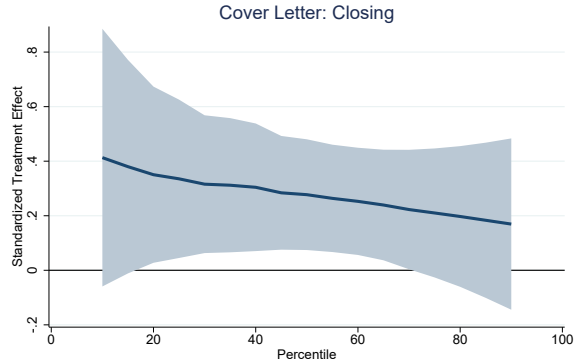


- Quantile regressions show that gains are concentrated at the **bottom of the distribution**.
- LLM access disproportionately benefits **lower-scoring applicants**.

Heterogeneity by section



Introduction



Closing

- The strongest gains appear in **standardized sections**.
- Effects decline across the distribution, consistent with **signal compression**.

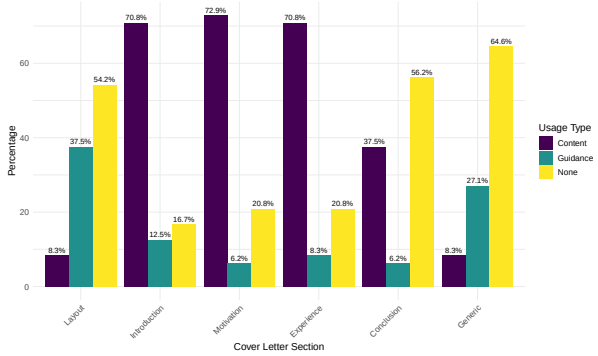
Clarity: two offsetting forces

	(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 mean	11.931	0.028	35.982	13.917	428.153	6.294
Observations	137	137	137	137	137	137

- ChatGPT reduces grammatical mistakes by **about 39%**.
- But it also makes text **harder to read** by increasing linguistic complexity.
- These offsetting effects explain the null result on recruiter-rated clarity.

Clarity table Readability figures

What do job-seekers use ChatGPT for?



- Job-seekers actively use ChatGPT for **motivation**, **experience**, and **introduction**.
- So the weak effects on personalized sections are **not** because applicants ignore them.
- They reflect a limitation of LLMs in **personalized persuasive writing**.

Prompt distribution

ChatGPT as substitute for search and other tools

- Treated job-seekers conduct:
  - **51% fewer Google searches**
  - **38% fewer website visits**
  - **84% fewer grammar-related searches**
- ChatGPT acts as a **bundled tool**:
  - writing assistance
  - company research
  - strategy advice
  - grammar correction

ChatGPT as substitute for search and other tools

- Treated job-seekers conduct:
  - **51% fewer Google searches**
  - **38% fewer website visits**
  - **84% fewer grammar-related searches**
- ChatGPT acts as a **bundled tool**:
  - writing assistance
  - company research
  - strategy advice
  - grammar correction
- So LLMs do not simply edit text, they **replace multiple stages** of the application process.

Browser history table Lee bounds

Experiment II: recruiter perceptions of AI

- Second experiment with **401 recruiters**.
- Each recruiter evaluates five pseudonymized cover letters drawn from Experiment I.
- Three information conditions:
  - **No Information**
  - **Partial Information**: some letters used ChatGPT
  - **Full Information**: exactly which letters used ChatGPT

Experiment II: recruiter perceptions of AI

- Second experiment with **401 recruiters**.
- Each recruiter evaluates five pseudonymized cover letters drawn from Experiment I.
- Three information conditions:
  - **No Information**
  - **Partial Information**: some letters used ChatGPT
  - **Full Information**: exactly which letters used ChatGPT
- Goal: identify whether recruiters can detect LLM use, and how **disclosure** changes evaluation.

Can recruiters detect LLM usage?

- In the partial-information treatment, recruiters are asked to guess which letters used ChatGPT.
- Correct detection rate is **49%**.
- Statistically indistinguishable from random guessing.

Can recruiters detect LLM usage?

- In the partial-information treatment, recruiters are asked to guess which letters used ChatGPT.
- Correct detection rate is **49%**.
- Statistically indistinguishable from random guessing.
- Interpretation:
  - recruiters know LLM use is widespread
  - but they cannot identify specific LLM-written texts reliably
- Therefore the relevant screening environment is one of **imperfectly observable adoption**.

Main recruiter-side results

	(1) Total	(2) Layout	(3) Intro	(4) Cover Letter Experience	(5) Motivation	(6) Closing	(7) Likelihood of Interview	(8) High Chance of Interview
Partial	-0.03	-0.10	-0.01	-0.02	-0.04	-0.05	-0.14	-0.16
	(0.09)	(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)
GPT-written	-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)
Full × GPT-written	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)

- Merely telling recruiters that some applicants used ChatGPT changes little.
- The interesting effects appear when **usage is explicitly disclosed**.

Recruiter main table

Disclosure does not penalize everyone equally

- Moving from **No Information** to **Partial Information**: essentially no effect.
- Moving to **Full Information**: modest average effects, but strong **heterogeneity**.

Disclosure does not penalize everyone equally

- Moving from **No Information** to **Partial Information**: essentially no effect.
- Moving to **Full Information**: modest average effects, but strong **heterogeneity**.
- Disclosure creates a **premium for strong letters that are known to be human-written**.
- This is most visible in the **upper tercile** of cover letters.

Upper tercile: recruiters reward high-quality human-written letters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Cover Letter				
	Total	Layout	Intro	Experience	Motivation	Closing	Likelihood of Interview	High Chance of Interview
Panel A. All Cover Letters (N=449)								
Full Info	0.18**	0.25***	0.15	0.20**	0.18**	0.16*	0.36**	0.49**
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.16)	(0.19)
Control Group Mean	0.14	0.11	0.11	0.15	0.09	0.09	3.56	0.55
*Panel B. Cover Letters Written **WITHOUT** ChatGPT Assistance (N=230)*								
Full Info	0.24**	0.36***	0.22*	0.22*	0.25**	0.26**	0.63***	0.82***
	(0.11)	(0.13)	(0.12)	(0.12)	(0.12)	(0.12)	(0.24)	(0.30)
Control Group Mean	0.27	0.11	0.19	0.28	0.18	0.26	3.67	0.60
*Panel C. Cover Letters Written **WITH** ChatGPT Assistance (N=219)*								
Full Info	0.12	0.15	0.08	0.18	0.11	0.06	0.09	0.20
	(0.13)	(0.12)	(0.14)	(0.13)	(0.13)	(0.13)	(0.22)	(0.26)
Control Group Mean	-0.01	0.11	0.02	0.02	0.01	-0.09	3.45	0.51

- At the top of the distribution, recruiters place more value on **high-quality letters written without ChatGPT**.
- So disclosure does not simply punish AI use; it **rewards clearly human-crafted strong signals**.

Upper-tercile table

Interpretation

- LLMs improve cover letters, but mostly in **generic and standardized** dimensions.
- They do not improve the dimensions recruiters care most about for interview selection.
- Gains are strongest for **weaker applicants**, which compresses the signal distribution.

Interpretation

- LLMs improve cover letters, but mostly in **generic and standardized** dimensions.
- They do not improve the dimensions recruiters care most about for interview selection.
- Gains are strongest for **weaker applicants**, which compresses the signal distribution.
- The assignment model in the paper formalizes this mechanism.
- Main implication: when signals are inflated and partially unobservable, **sorting becomes less efficient**.
- Calibration implies **welfare losses through mismatch**.

Conclusion

- LLMs improve the **average quality** of labor market signals.
- But the gains are concentrated in **less personalized and less decision-relevant** parts of the cover letter.
- Therefore they do **not** increase interview recommendations in our setting.

Conclusion

- LLMs improve the **average quality** of labor market signals.
- But the gains are concentrated in **less personalized and less decision-relevant** parts of the cover letter.
- Therefore they do **not** increase interview recommendations in our setting.
- More broadly, LLMs **compress** the distribution of written signals and weaken their informativeness about applicant ability.
- When recruiters know whether LLMs were used, they place more value on **high-quality human-written** letters.
- The broader implication is that LLMs can distort matching even when they improve text quality on average.

Bottom line: LLMs raise average signal quality, but reduce signal informativeness.

▸ Appendix menu

Appendix

Main deck appendices

- Main results table
- Interview determinants
- Clarity table
- Prompt distribution
- Browser history
- Lee bounds

Recruiter / model appendices

- Recruiter main table
- Upper-tercile table
- Readability figures
- Model summary
- Identification at a glance
- Clarity vs signaling

Main deck

Appendix: Main results table

	Total	Layout	Intro	Exp.	Motiv.	Closing	Int. like.	High int.
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)
Obs.	274	274	274	274	274	274	274	274

Appendix menu Main deck

Appendix: Determinants of interview likelihood

	Likelihood of interview
CL: Layout	0.525**
	(0.216)
CL: Introduction	0.293
	(0.249)
CL: Experience	0.399
	(0.249)
CL: Motivation	0.823***
	(0.249)
CL: Closing	0.474
	(0.315)

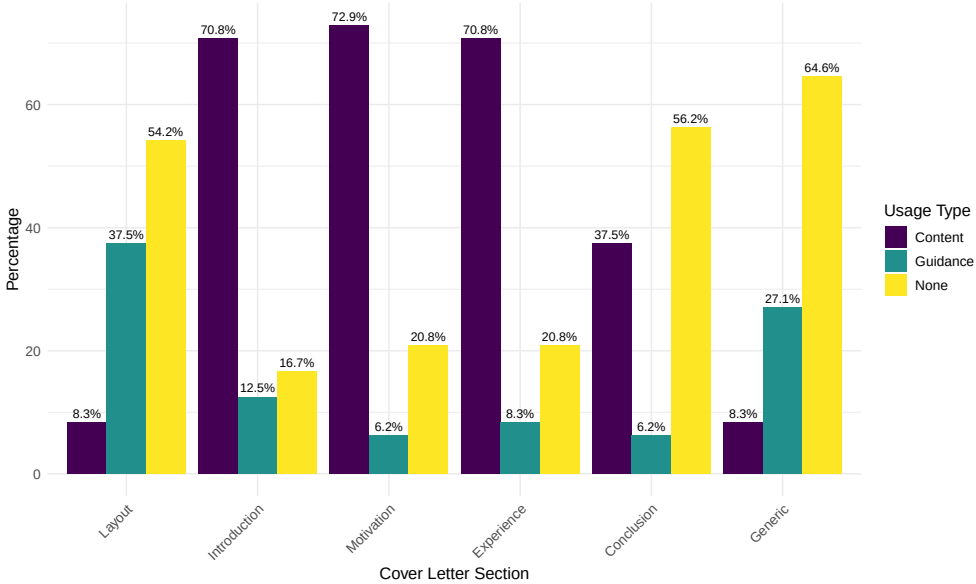
Appendix menu Main deck

Appendix: Clarity and readability table

	Errors	Error rate	Flesch ease	FK grade	Words	Avg 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 mean	11.931	0.028	35.982	13.917	428.153	6.294
Obs.	137	137	137	137	137	137

Appendix menu Main deck

Appendix: Prompt distribution



Appendix: Browser history

	# Websites	# Google	Firm	CL	Grammar	Translation
ChatGPT	-7.422***	-5.120***	-1.825	-1.140	-2.427***	-1.280
	(2.215)	(1.320)	(1.169)	(0.692)	(0.631)	(0.749)
Control mean	19.351	10.000	5.281	2.368	2.895	2.175
Obs.	119	119	119	119	119	119

Appendix menu Main deck

Appendix: Browser-history Lee bounds

Upper bound and lower bound both keep the qualitative result: treated applicants browse less.

	# Websites visited	# Google searches
Lee upper bound	-2.220	-1.939*
Lee lower bound	-9.484***	-6.341***

Appendix menu Main deck

Appendix: Recruiter experiment main table

	Total grade	Layout	Interview likelihood	High interview chance
Partial	-0.03	-0.10	-0.14	-0.16
Full	0.06	0.08	0.19	0.26*
Written with GPT	-0.04	0.03	-0.10	-0.09
Partial × GPT	0.04	0.14	0.14	0.14
Full × GPT	0.02	0.02	-0.15	-0.25

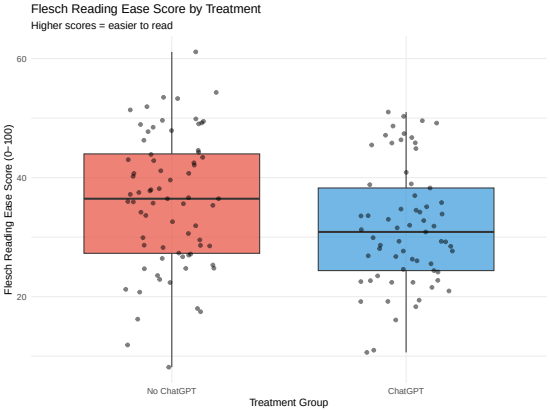
Appendix menu Main deck

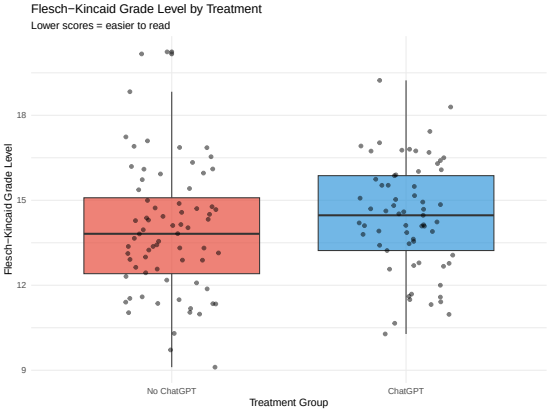
Appendix: Upper-tercile recruiter result

	Total grade	Motivation	High interview chance
Partial	-0.16	-0.20*	-0.56*
Full	0.09	0.06	0.26
Written with GPT	-0.48***	-0.48***	-0.90***
Partial × GPT	0.20	0.31*	0.54
Full × GPT	0.04	0.12	-0.08

Appendix menu Main deck

Appendix: Readability figures

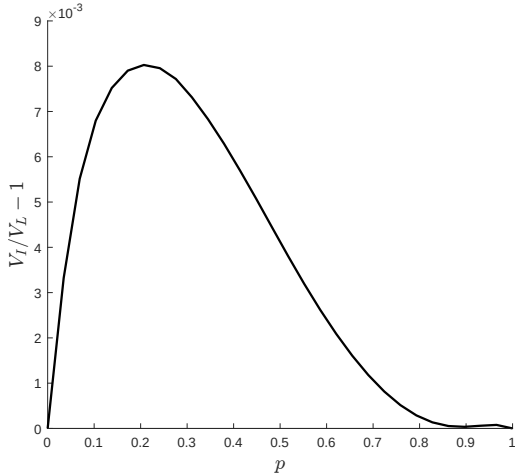




▸ Appendix menu ◂ Main deck

Appendix: Model summary

- Assignment model with imperfect information.
- LLMs disproportionately improve signals at the bottom.
- Firms cannot perfectly observe LLM adoption.
- This compresses signals and worsens sorting.
- Calibrated efficiency losses can reach **0.8%**.



Appendix menu Main deck

Backup: identification at a glance

- **Experiment I**: random assignment of ChatGPT access identifies the causal effect of LLM access on cover letter quality.
- Recruiters are blind to treatment status in the first experiment.
- **Experiment II**: randomizes disclosure, identifying how recruiter beliefs about LLM usage affect evaluations.
- Together the two experiments separate:
  - the **effect of LLM use on the signal**
  - the **effect of knowing about LLM use on evaluation**

Appendix menu Main deck

Backup: clarity versus signaling

- **Clarity channel**: better writing reduces information frictions.
- **Signaling channel**: better writing is itself informative about worker ability.

Backup: clarity versus signaling

- **Clarity channel**: better writing reduces information frictions.
- **Signaling channel**: better writing is itself informative about worker ability.
- In our setting:
  - ChatGPT improves spelling and grammar
  - but worsens readability
  - and disproportionately boosts weaker applicants
- Therefore LLMs do not simply improve clarity; they also **change the informational content of the signal**.

Appendix menu Main deck