‘Do AI skills improve employment chances? Evidence from a field experiment’

Anastasia Danilov, Humboldt-Universität zu Berlin

Teo Firpo, Humboldt-Universität zu Berlin

Lukas Niemann, without academic affiliation

**** This version: Augus 30, 2024 ****

**** Please do not cite or distribute ****

With the advent of Artificial Intelligence (AI), a growing number of firms are adopting this technology, leading to a rise in demand for AI-related skills. There is, however, little evidence on the impact of complementary AI skills for prospective job seekers in roles where AI skills are not the main focus. We address this question with a correspondence study. We send résumés, randomizing whether they contain AI skills, to 1,185 entry-level job ads in the UK across a range of job functions. We find no evidence that including complementary AI skills in a résumé increases a candidate’s chances of being invited for an interview. We explore heterogeneity of our results by the type of vacancy, finding that AI skills are valued by employers advertising more specialized job roles. In a survey of professionals with hiring experience, we test the robustness of our results.

Keywords: AI skills, hiring, investment in human capital, field experiment

JEL classification: O33, J23, J24, I26

1. Introduction

Artificial Intelligence (AI) is seen as the next 'General Purpose Technology' (Brynjolfsson et al., 2019; Goldfarb et al., 2023). Indeed, in the past decade firms around the world have increasingly begun adopting AI tools (Babina et al., 2024). The surge in AI technology adoption has subsequently elevated the demand for AI expertise, resulting in a significant wage premium for employees possessing AI skills (Alekseeva et al., 2021; Squicciarini and Nachtigall, 2021; Stapleton et al., 2021; Zhang et al., 2021).

This trend implies that job candidates with AI skills should experience enhanced employment prospects. However, to the best of our knowledge, there is no empirical evidence assessing the impact of candidates' AI skills on hiring prospects.

This paper provides causal empirical evidence on the return to AI skills for job candidates seeking positions of young professionals. In a randomized controlled trial we submit 1,185 online job applications to open entry-level vacancies across a range of job types available on job-search platforms in the United Kingdom, randomizing whether the candidates' résumés include AI skills or not. Overall, we find that listing AI skills in résumés did not improve the chances of being invited for a job interview. Nevertheless, we observe some heterogeneity by job function, with AI skills having a positive effect on application success for marketing and engineering roles, but with no positive effects for HR, finance, logistics and IT positions. When analyzing the text content of the job ads to which the applications are sent, we find no connection between mentioned digital skills in the job postings and our treatment effects. Moreover, we analyze the skills mentioned in each job vacancy using the ESCO taxonomy, to explore heterogeneity by the degree of specialization of advertised roles. We use the number of ESCO skill categories mentioned in the job ad as a proxy for specialization, with fewer categories indicating more specialized job roles, and more categories indicating more generalist roles. We find that listing AI skills has a positive and statistically significant impact on callbacks when applying to vacancies mentioning fewer skill categories, suggesting that AI skills may be more beneficial for highly specialized roles. (These results are not included in this draft yet.) These results may reflect a higher value of AI skills in the type of occupations that are most exposed to the risk of automation from AI, as has been argued in recent literature (Acemoglu et al., 2020; Webb, 2020).

Our findings present a surprising picture, suggesting that employers on average did not value AI skills of job candidates in their recruitment process. We explore a number of explanations for our findings: (i) whether recruiters notice the AI skills in our résumés; (ii) whether the AI skills we used are perceived as relevant; (iii) whether these skills signal higher general competencies; (iv) and whether candidates with the AI skills are perceived as overqualified (and therefore might be not a good fit). To do so, we survey 771 professionals with hiring experience by showing them résumés (with and without AI skills) in random order from a job function of their expertise. We find that AI skills in our résumés are noticeable and considered as relevant for the job functions that they target; that the listed AI skills also signal higher digital and cognitive skills of applicants (but lower social skills); and that candidates with AI skills are marginally more likely to be considered as overqualified. The survey participants expect the candidates with AI skills to receive more invitations for interviews.

Our paper contributes to the literature on the return to skills. Grasping the return on investment in AI skills and their value is pivotal for understanding the training decisions of individuals and firms, as suggested by Gary Becker's theory of investment in human capital (Becker, 1964).

There are many reasons for firms to seek workers with AI skills. AI is associated with increases in productivity (Czarnitzki et al., 2023). Moreover, the literature on human-machine interactions provides strong evidence for the complementarity of skills related to AI technology with other job-relevant skills across a wide range of jobs (Chugunova and Sele, 2022). Indeed, recent studies have shown that using AI skills can increase productivity across a range of roles from consulting (Dell'Acqua et al., 2023) to customer services (Brynjolfsson et al., 2023). Taken together, this evidence suggests that employers should value candidates' AI expertise, as these competencies are likely to make them and the firm more productive[1], even if the vacancy they are aiming to fill is not directly related to AI.

The rest of the paper is structured as follows. In Section 2, we describe the research methodology. Section 3 outlines the results. Section 4 discusses our findings and concludes.

[1] Because skills related to AI are likely to be general rather than firm-specific,
AI is understood to be a General Purpose Technology (Brynjolfsson et al., 2019; Goldfarb et al., 2023).

2. Research methodology

To understand the effect of AI skills on the success of the job search, we conduct a field experiment employing the 'correspondence study' methodology that has been largely used to study labor market discrimination (for a review, see Riach and Rich, 2002), ethnical discrimination (Bartoš et al., 2016), as well as résumé characteristics such as qualifications (Verhaest et al., 2018) or entrepreneurial background (Kacperczyk and Younkin, 2022). We send job applications to 1,185 job search adverts published by UK[2] companies in the months March through November 2021. We randomly assign job adverts (one per company) to one of two experimental conditions: those that receive résumés with AI competencies (henceforth 'treatment group'), and those that receive résumés without any mention of such skills ('control group'). In total, we send 591 and 594 résumés to control and treatment groups, respectively.

We send résumés for roles across six job functions. We select job functions to cover a variety of typical positions inside firms, spanning operational, production and sales roles, including: Finance and Accounting; Human Resources (HR); Marketing; Information Technology (IT); Logistics and Supply Chains; and Engineering. These functions reflect the growing opportunities to apply AI solutions within firms, with an increase in demand for AI in IT, professional services, finance and manufacturing functions (Alekseeva et al., 2021; Squicciarini and Nachtigall, 2021; Stapleton et al., 2021).

For each job function, we create two identical résumés that only differ in their inclusion of AI skills. Treated résumés include: a brief mentioning of AI skills in the short 'Personal statement' text at the top of the résumé; one extra item in the 'Skills' section mentioning competencies such as 'IBM Watson' or 'Machine Learning'; an additional bullet point with AI skills in both the 'Education' and 'Work Experience' sections; and a record of an AI online course under 'Additional Experience'. These skills are selected based on recent case studies and data on the most frequently demanded AI skills from Alekseeva et al. (2021). All other details regarding the candidate's training and experience remain exactly the same. Our

[2] Unlike in other countries (e.g., Germany), in the UK labor market during the time when study was conducted it was common to send job applications consisting only of résumés and a motivation letter, without supplementary documents, such as transcripts, references, or certifications.

candidate has a master's degree in the respective job function and one-year job experience. All résumés can be found in the Online Appendix[3].

We construct our sample of job ads by searching for open positions using UK online job boards such as Indeed.co.uk, Monster.co.uk, and Reed.co.uk[4]. We restrict our search to job offers that require either no, or little professional experience (up to one year). We avoid job openings that ask for specific, uncommon skills (e.g., speaking Mandarin). We follow a predefined random allocation to decide whether to send the résumé that includes AI-skills (treatment) or not (control). To avoid spill-over effects, we do not apply for more than one job per company. The distribution of applications by job function is reported in Table 1 below.

We explicitly focus on the job functions that do not work directly on developing AI solutions, such as software development or scientific research. First, our main interest is on the complementary effect of AI skills to other job skills and experiences. Second, jobs in AI development usually require more comprehensive documentation on the applicant's track record (references to GitHub account, scientific publications) which cannot be easily or ethically faked. In this respect, our study should be interpreted as providing evidence for the returns to AI skills across a broader range of job roles where this technology can be deployed and implemented, rather than outright developed.

[3] The Online Appendix can be found at the following link: [https://www.dropbox.com/scl/fo/bwthe3672hlh9wy81eqs4/AK7E3-51KouVfQ7cKbvbpAQ?rlkey=pi8n58y0jd655hhjgqmbn070g&st=xj150dtq&dl=0] *<temporary link, the repository link will be added after the review>*

[4] The full list of job boards used includes: 'CV Library', 'Indeed', 'Milkround', 'Monster', 'Reed', 'Totaljobs'.

Table 1 – Applications by job function and treatment group

Job function	Control	Treatment	Total
Finance and accounting	94	95	189
HR	97	99	196
Marketing	100	100	200
IT	100	100	200
Engineering	100	100	200
Logistics and supply chains	100	100	200
Total	591	594	1,185

We focus on two measures of 'callback' rates[5]. The first is a strict measure that only includes explicit invitations for an interview (henceforth 'strict callback'). The second is a broader measure that incorporates all positive responses including or in addition to interview invitations, e.g., offers to interview for other roles, requests to call back, etc. (henceforth 'broad callback'). We collect responses over the course of 40 days after the application. To minimize any inconvenience on the part of the employers, if the candidate receives a callback we reply in a timely manner with a brief and polite rejection.

Additionally, we collect data on the text description of the job roles from the job ads, extracting the skills requested from the applicants[6].

The experiment was pre-registered in the AEA Registry with reference number <*will be added after the review*>. We obtained ethical approval from the German Association for Experimental Economic Research e.V. (<*will be added after the review*>).

3. Results

We begin with descriptive statistics of the results. Table 2 reports the callback rates (i.e. including treatment and control groups) by job function. In Panel A, we report the strict

[5] These measures follow from the literature on audit experiments (Baert et al., 2015; Verhaest et al., 2018).

[6] These data could not be collected for 10 ads overall (less than 1% of the sample).

callback rates. In Panel B, we report the broad callback rates. Callback rates range between 3.17% and 11.5%, for the strict callback, and 6.35% and 19.50% for the broad callback, being slightly lower for the function Finance and Accounting than for other job functions (p-values < 0.1, see a Probit regression in Table A1 of the Appendix for an additional statistical analysis).

When it comes to the effects of the AI treatment, we find no overall effect, as the callback rates between treatment and control groups are not significantly different ($p = 0.407$, *test statistic*= -9 for strict callback rates; and $p = 0.201$, test statistic = -17 for broad callback rates; two-sided Fisher-Pitman permutation test with 200,000 runs, henceforth 'FPP test'). This masks some heterogeneity at the job function level. In functions 'Marketing' and 'Engineering', the callback rates in the treatment group are at least twice as high as in the control group. (For 'Marketing', 16.00% vs. 7.00% for strict callback rates and 24.00% vs. 12.00% for broad callback rates; for 'Engineering', 10.00% vs. 4.00% strict callback rates and 20.00% vs. 8.00% for broad callbacks rates). The differences in 'Marketing' are statistically significant (marginally for strict callback rates, with $p = 0.075$, test statistic = -9; for broad callback rates, with $p = 0.042$, test statistic = -12; FPP tests). For 'Engineering', the difference in broad callback rates is significant ($p = 0.023$, test statistic = -12) but not the difference in strict callback rates ($p = 0.163$, test statistic = *-6*; FPP tests).

Table 2 – Callback rates by job function and treatment

Panel (A): Callback rates in strict sense

Job function	Overall	AI	Control	Ratio (AI/control)	*p*-values	Test statistic
Finance	3.17%	3.16%	3.19%	0.99	1	0
HR	7.65%	9.09%	6.19%	1.47	0.621	-3
Marketing	11.50%	16.00%	7.00%	2.29	0.075*	-9
IT	7.50%	5.00%	10.00%	0.50	0.283	5
Engineering	7.00%	10.00%	4.00%	2.50	0.163	-6
Logistics	11.00%	9.00%	13.00%	0.69	0.502	4
All functions	8.02%	8.75%	7.28%	1.20	0.403	-9

Panel (B): Callback rates in broad sense

Job function	Overall	AI	Control	Ratio (AI/control)	*p*-values	Test statistic
Finance	6.35%	6.32%	6.38%	0.99	1	0
HR	13.78%	12.12%	15.46%	0.78	0.637	3
Marketing	18.00%	24.00%	12.00%	2.00	0.042**	-12
IT	16.50%	13.00%	20.00%	0.65	0.254	7
Engineering	14.00%	20.00%	8.00%	2.50	0.023**	-12
Logistics	19.50%	21.00%	18.00%	1.17	0.725	-3
All functions	14.77%	16.16%	13.37%	1.21	0.203	-17

Note: Callback rate in the 'strict sense' only includes direct invitations to interview. Callback rate in the 'broad sense' includes all positive responses from employers (e.g., request to call back, interview for another post). *p*-values and test statistics for the difference between the AI treatment and control from a two-sided Fisher-Pitman permutation test with 200,000 runs. N = 1,185. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The regression analysis reported in Table 3 confirms the results presented above. We use the linear probability model for ease of interpretation[7]. The dependent variable is a dummy variable equal to one if the application received a callback, and zero otherwise. Models (1) and (4) present the estimation of the average treatment effect for the strict and broad callbacks respectively. Models (2) and (5) include dummies for each of the job functions. Models (3) and (6) add interaction terms for the AI treatment and each of the job function dummies. Across all models, the effect of having AI skills on the likelihood of receiving a callback is not statistically significant (*p*-values range from 0.172 to 0.990). However, the interaction terms of the

[7] Our pre-analysis plan specified both linear and Probit regressions. However, since the results from the two estimation methods do not differ (either in the direction, size, or statistical significance of the coefficients), we present here only the former.

treatment dummy and the dummy for 'Marketing' are positive and marginally significant at the 10% level for the strict (p = 0.082) and broad callback (p = 0.064). This suggests that the likelihood to receive a callback for a marketing job is significantly higher when a candidate lists AI skills in their résumé (9 percentage points for a strict callback and 12 percentage points for a broad callback). The interaction term of the treatment dummy and the 'Engineering' dummy is also positive and statistically significant but only for the broad callback (12.1 percentage points, p = 0.047 for the broad callback). Controlling for whether the ad directly mentioned any of the skills our candidate highlighted in the résumé does not change our results (see Table A2 in Appendix).

The sample size and statistical power should also be considered when interpreting the result. The sample size in the correspondence study (1,185) allows us to detect an effect size of 4.2 percentage points or more for the strict callback rate, if we assume power of 0.80 and $\boldsymbol{\alpha}$ = 0.05, and given the observed mean strict callback rate of 0.073 and the standard deviation of 0.260 for the control group. For the broad callback rate this effect size would amount to 5.6 percentage points (given the mean broad callback rate of 0.134 and the standard deviation of 0.341 in the control group).[8] As a result, our findings should be understood as an upper bound for the effect of having AI skills on a job seeker's likelihood to receive a callback rate, since we cannot exclude that the real effects were positive but smaller than the minimum detectable effect sizes reported above.

[8] For the power analysis we use Stata's *power* command.

Table 3 – Regression results

	(1)	(2)	(3)	(4)	(5)	(6)
		Strict callback			Broad callback	
Treatment	0.015	0.015	-0.000	0.028	0.028	-0.001
	(0.016)	(0.016)	(0.026)	(0.021)	(0.021)	(0.036)
HR		0.045*	0.030		0.074**	0.091**
		(0.023)	(0.031)		(0.031)	(0.045)
Marketing		0.083***	0.038		0.117***	0.056
		(0.026)	(0.032)		(0.033)	(0.042)
IT		0.043*	0.068*		0.102***	0.136***
		(0.023)	(0.035)		(0.032)	(0.048)
Engineering		0.038*	0.008		0.077**	0.016
		(0.022)	(0.027)		(0.030)	(0.037)
Logistics		0.078***	0.098**		0.132***	0.116**
		(0.026)	(0.039)		(0.033)	(0.046)
Treatment *x* HR			0.029			-0.033
			(0.046)			(0.061)
Treatment *x* Marketing			0.090*			0.121*
			(0.052)			(0.065)
Treatment *x* IT			-0.050			-0.069
			(0.045)			(0.064)
Treatment *x* Engineering			0.060			0.121**
			(0.044)			(0.061)
Treatment *x* Logistics			-0.040			0.031
			(0.052)			(0.067)
Observations	1,185	1,185	1,185	1,185	1,185	1,185
Adj. R-squared	-0.000	0.001	0.011	0.001	0.011	0.017

Note: Linear probability model (OLS) with robust standard errors. The dependent variable 'callback' is equal to one if the application received a call back, and zero otherwise. The variable 'Treatment' is equal to one in the treatment AI and zero otherwise. The variables 'HR', 'Marketing', 'IT', 'Logistics' and 'Engineering' are dummy variables equal to one for the specific job function and zero otherwise. The function Finance is the reference category. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Overall, our findings present a mixed picture on the returns to AI skills. While there seem to be some significant advantages for applicants with AI skills to entry-level marketing and engineering roles, listing AI skills in résumés has no significant effect for applications in other functions. These results are puzzling in light of the wider evidence on the demand for AI-skills: while 'Marketing' and 'Engineering' have been identified as areas with high potential use for AI solutions, the same is true for 'Logistics and Supply Chains' and 'IT', with the remaining categories not far behind (Chui et al., 2018). Given this potential for applying AI competencies across job functions, one would expect a starker difference in callback rates between treatment and control groups.

To ensure that our observations are not driven by the design of our résumés and choice of AI skills, we explore whether (i) AI skills are easy to notice in the résumés; (ii) AI skills are relevant for the types of job roles we apply for with our résumés; (iii) the résumés with AI skills signal higher general competencies; (iv) the candidates represented by résumés with AI skills are perceived as overqualified. To do so we conduct a survey of professionals with professional work experience in the six job functions we study, and experience in hiring. Using Prolific.com, we recruit a total of 771 participants for an online survey. The survey consists of three parts. In Parts I and II, the participants are presented with the résumés that we used in the correspondence study. In Part I, participants are randomized to see either the treatment or control résumé from the job function relevant to their professional experience (e.g. the 'Marketing' résumé for professionals with experience in marketing). They are then asked questions about the candidate's skills and chances of getting invited for a job interview. In Part II, they are shown the other version of the résumé, and asked the same questions. In Part III, participants are asked more questions about both résumés and general questions about the value of AI skills in the labor market. The survey can be found in the Online Appendix.

First, we explore whether the participants notice the presence of AI skills in the résumés. Attention paid to résumé characteristics might be scarce. If the survey participants do not notice the AI skills, it is likely that the actual recruiters in the correspondence study did not notice them either, which could explain the null results reported above. To test this, we ask participants in Part III of the survey to state which of the two résumés they saw mentioned AI competencies. Respondents do not have the option to go back and look at the résumés again. We incentivize correct answers with an additional payment of £0.10. We find that 95.46% of participants can recall the order number of the résumé containing the AI skills correctly. The difference with a random choice of 50% is statistically significant (*$p = 0.000$*; tested with a two-sided Binomial probability test). The probability of recalling the order correctly is similar across all job functions (see Table A3 in the Appendix).

Second, we explore whether the participants perceive the AI skills mentioned in the résumés as relevant for the role for which the candidate applied. If they are not perceived as relevant, it could explain why AI résumés do not receive higher callback rates. Recruiters in the correspondence study may have judged, despite seeing the AI skills, that they would not affect the candidate's ability to perform the tasks required for the role. To test this, in Part III we ask participants to rate their agreement with the statement "The AI skills [*list of AI skills from the résumé*][9] would significantly improve a candidate's ability to execute the [*job function] assistant role*" on a scale from 1, 'Strongly disagree' to 5 'Strongly agree'. We find that 57.07% of respondents select 'Agree' or 'Strongly agree' across functions, with only 19.33% selecting 'Disagree' or 'Strongly disagree'. The difference is statistically significant (two-sided one sample test of proportions, $p = 0.000$; *z-score* = 26.537). Figure A1 in the Appendix shows the distribution of responses by job function. The percentage of participants agreeing with the above statement is higher than the percentage of those disagreeing across all job functions, and the differences are statistically significant (all $p < 0.000$ with two-sided one sample tests of proportions)[10].

Third, we explore whether the AI skills signal higher ability. As suggested by signaling theory, more able individuals tend to invest in more education (Spence, 1973; Arrow, 1973). The acquired education might not necessarily have a direct effect on their productivity, but act as a costly sorting device that enables employers to identify more able individuals. As long as employers hold such expectations, even if the specific skill set is not applicable to the job role, one would expect employers to select candidates with stronger signals of ability and cognitive function. We test whether this is the case in Parts I and II by asking respondents to evaluate the competencies of the candidate represented in the résumé, asking them "How would you assess the candidate's [*skill*]?" on a Likert scale from 1 'Very low' to 5 'Very high' for five types of

[9] These include examples from the relevant résumé. (E.g. participants who see the Marketing résumé are shown examples of AI skills present in the Marketing résumé).

[10] The exact differences are reported in Table A4 in the Appendix.

skills. We do this for both the résumés and report the within-subject means for each skill in Table 4 below.

Table 4 – Perceived candidate skills by résumé (Prolific survey)

	AI résumé	Control résumé	P-value	Test statistic
AI skills	4.125	2.523	0.000***	1235
Advanced digital skills	4.014	3.482	0.000***	401
Basic digital	4.454	4.187	0.000***	206
Cognitive skills	4.014	3.844	0.000***	131
Social skills	3.744	3.808	0.009***	-49

Note: Results from Prolific study data. Each row presents the mean response from 1 'Very low' to 5 'Very high' to the question on the corresponding skill for the relevant résumé. P-values and test statistics from a two-sided Fisher-Pitman permutation test of paired replicates with 200,000 runs. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. N = 771.

Our overall finding is that participants rate skills of the candidates in the treatment group more highly, with the largest difference for AI skills (all differences are statistically significant with all $p < 0.01$). The only exception is social skills for which the reverse is true, with the AI résumés scoring 3.744 on average and the control résumés 3.808 (the difference being statistically significant with $p = 0.010$; test statistic = -49, FPP test of paired replicates).

Fourth, we test whether candidates with résumés containing AI skills are perceived to be overqualified. Perceptions of under- or overqualification may have two opposite effects. On the one hand, underqualified candidates may be rejected for lacking the necessary skills. On the other hand, recruiters may perceive overqualification as a negative signal (for instance, because a more qualified candidate may demand higher wages or be more likely to leave soon for another role). To test this, we ask participants "In your opinion, how well do the candidate's qualification and experience fit an entry-level [*job function*] position?" on a Likert scale from 1 'Underqualified' to 5 'Overqualified' (with 3 being 'Neither under- nor overqualified'). We again do this for both the résumés, and report the within-subject means. Indeed, we find some evidence that respondents are more likely to consider candidates with AI skills as overqualified, with a mean response of 3.838 (SD: 0.763) against 3.790 (SD: 0.749) in the control. The difference is marginally statistically significant ($p = 0.097$, test statistic = 37.0; FPP test of paired replicates). Table A5 in the Appendix reports the differences by job function. Across all functions, except for Engineering, the candidate with AI skills is more likely to be considered overqualified, but magnitudes vary and none of the job function-level differences are statistically significant (all $p > 0.131$).

Fifth, we investigate if participants anticipate candidates featuring AI skills on their résumés would receive more interview invitations compared to those without such skills. To test this, in Parts I and II we inform participants that the résumé they saw was sent to real recruiters in the correspondence study, and ask them to estimate how many times out of 50 the candidate was invited for an interview. We incentivize correct answers with a bonus payment of £0.10. Participants are told they will receive the bonus payment as long as their guess deviates from the correct number by less than 3 integers.[11] We repeat this question for both treated and control résumés and report the within-subject results. For the résumés containing the AI skills, the participants estimate on average that the number of invitations was 31.45 out of 50 (SD: 13.30), whilst the average estimate for the control résumés stands at 28.69 (SD: 13.03). The difference is statistically significant ($p = 0.001$; test statistic = 2124; FFP test). For all job functions, the survey participants estimated that the résumé with the AI skills received more invitations, and the difference is statistically significant (all $p < 0.05$; FPP test of paired replicates. See Table A6 in the Appendix for more details).

4. Discussion and conclusion

This paper provides empirical evidence for the effect of including AI skills in a job seeker's résumé on their likelihood of being invited for a job interview for an entry-level position. Our main finding — that résumés with AI skills do not receive more callbacks from six different organizational functions compared to résumés without such skills — presents a surprising picture.

[11] To set the payments, we rescale the estimates according to the actual number of applications sent in that specific job function and treatment condition.

Other studies show that there is a growing demand for AI experts (Alekseeva et al., 2021; Stapleton et al., 2021) and that both the amount of AI-related job postings and the amount of AI skills within a single job posting have increased over time (Squicciarini and Nachtigall, 2021). Moreover, the empirical literature shows a growing number of applications for AI competencies across professional functions (Chui et al., 2018). Given these facts one would expect the résumés with AI skills to receive a higher number of job interview invitations compared to control résumés. But this is not what we observe. Nevertheless, professionals with hiring experience in our Prolific survey expect résumés with AI skills to receive more callbacks than control résumés. Results from the Prolific study show that participants notice the AI skills in our résumés and consider them relevant for the type of roles for which we apply. Furthermore, we find that the candidates represented by résumés with AI skills are considered to be more skilled across four dimensions, including AI, cognitive and digital skills, but to have slightly lower social skills.

There are three limitations of our study. First, the statistical power of our sample size. Although our sample size of 1,185 observations allows us to identify the minimum effect size of 4.2 percentage points with a power of 0.80, we lack the power to be able to identify effect sizes below this value. The lack of an overall significant difference between the treatment and control group suggests that adding AI skills to a résumé does not increase the likelihood of the callbacks by this or any higher number.

Second, our experiment allows us to track the hiring process only up to (at most) an invitation to an interview. As a result, we are limited in the extent to which we can draw any conclusions for the hiring process overall. As many previous correspondence studies have argued, however, a lower interview rate is likely to translate into lower expected hiring for candidates (Bertrand and Mullainathan, 2004). Relatedly, we do not observe wage offers to each job. Therefore, we cannot exclude that candidates with AI skills are offered higher wages, even if their chances of being invited to an interview and eventually hired are similar to the candidates without AI skills.

Third, we are limited in the range of job roles to which we can apply. This is because more senior positions require more specialized professional experience, which cannot be met by the identically held résumés. This poses a problem since the skills we focus on – AI – might prove to be more suitable only in more experienced roles. We partially address this question in our Prolific study, by asking respondents whether the skills provided in the résumés are relevant for the types of the job positions we apply for in our main study. While our findings support the hypothesis that the skills are relevant for the entry-level positions we apply for, we cannot exclude that the effect of including AI skills in résumés of more senior job seekers could lead to higher callback rates.

Our results indicate that including AI skills in the résumés does not have a strong positive effect on the actual likelihood of invitation to a job interview. At the same time, we show that job candidates with AI skills are perceived as more qualified and skilled for entry-level positions where AI skills are mostly complementary. These findings demonstrate a gap in the perception of AI skills and whether they are valued by employers. Future research could focus on how this trend will develop over time.

References

Acemoglu, D., Autor, D., Hazell, J., Restrepo, P. (2022). Artificial Intelligence and Jobs: Evidence from Online Vacancies. *Journal of Labor Economics*, *40*(S1), S293–S340.

Alekseeva, L., Azar, J., Gine, M., Samila, S., Taska, B. (2021). The demand for AI skills in the labor market. *Labour economics, 71*, 102002.

Arrow, K. J. (1973). The Theory of Discrimination. In O. Ashenfelter, A. Rees (Ed.), *Discrimination in Labor Markets* (pp. 3–33). Princeton University Press.

Babina, T., Fedyk, A., He, A., Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, *151*, 103745.

Baert, S., Cockx, B., Gheyle, N., Vandamme, C. (2015). Is there less discrimination in occupations where recruitment is difficult? *Industrial and Labor Relations Review*, *68*(1), 467–500.

Bartoš, V., Bauer, M., Chytilová, J., Matějka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, *106*(6), 1437-75.

Becker, G. S. (1964). *Human Capital* (2nd ed.). New York: Columbia University Press.

Bertrand M., Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review* *94*(1), 991–1013.

Brynjolfsson, E., Li, D., Raymond, L. (2023). Generative AI at Work. NBER Working Paper Nr. 31161.

Brynjolfsson, E., Rock, D., Syverson, C. (2019). 1. Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. In A. Agrawal, J. Gans, A. Goldfarb (Ed.), *The Economics of Artificial Intelligence: An Agenda* (pp. 23–60). Chicago: University of Chicago Press.

Chugunova M., Sele, D. (2022). We and It: An interdisciplinary review of the experimental evidence on how humans interact with machines. *Journal of Behavioral and Experimental Economics, 99*, 101897.

Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., Malhotra, S. (2018). Notes from the AI frontier: Insights from hundreds of use cases. *McKinsey Global Institute*, *2*, 267.

Czarnitzki D., Fernández G. P., Rammer C. (2023). Artificial Intelligence and firm-level productivity. *Journal of Economic Behavior and Organization 211*, 188–205.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, 24–013.

Goldfarb, A., Taska, B., Teodoridis, F. (2023). Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings. *Research Policy, 52*, 104653.

Kacperczyk, A., Younkin, P. (2022). A Founding Penalty: Evidence from an Audit Study on Gender, Entrepreneurship, and Future Employment. *Organization Science*, *33*(2), 716–745.

Riach, P. A., Rich, J. (2002). Field experiments of discrimination in the market place. *The Economic Journal*, *112*(483), F480–F518.

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

Squicciarini, M., Nachtigall, H. (2021). Demand for AI skills in jobs: Evidence from online job postings. *OECD Science, Technology and Industry Working Papers*, 2021(3), 1–74.

Stapleton, K., Copestake, A., Pople, A. (2021). AI, firms and wages: Evidence from India. *SSRN Electronic Journal*.

Verhaest, D., Bogaert, E., Dereymaeker, J., Mestdagh, L., Baert, S. (2018): Do Employers Prefer Overqualified Graduates? A Field Experiment. *Industrial Relations, 57*, 361–388.

Webb, M. (2020). The Impact of Artificial Intelligence on the Labor Market. *Working paper*. https://www.michaelwebb.co/webb_ai.

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J. C., Sellitto, M., Shoham, Y., Clark, J., Perrault, R. (2021). The AI Index 2021 Annual Report. *AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, March 2021.*

Supplementary appendix – Additional tables and figures

Table A1 – Overall callback rates by job function (Probit regressions)

Job function	(1) Strict	(2) Broad
HR	0.427*	0.436**
	(0.222)	(0.181)
Marketing	0.655***	0.611***
	(0.213)	(0.176)
IT	0.416*	0.552***
	(0.222)	(0.177)
Engineering	0.380*	0.446***
	(0.224)	(0.180)
Logistics	0.629***	0.666***
	(0.214)	(0.175)
Observations	1,185	1,185

Note: Results from a Probit regression with the correspondence study data. The dependent variables are dummies equal to one for strict callback in column (1) and for broad call back in column (2), and zero otherwise. The independent variables are dummies for job functions with 'Finance and accounting' in the reference category. The names of job functions are shortened for clarity but follow the same order as Table 1 in the main text. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A2 – Regression results (LPM) with additional controls

	(1)	(2)	(3)	(4)
	Strict callback		Broad callback	
Treatment	0.015	0.000	0.028	-0.000
	(0.016)	(0.026)	(0.021)	(0.036)
Treatment *x* HR		0.029		-0.033
		(0.046)		(0.061)
Treatment *x* Marketing		0.091*		0.121*
		(0.052)		(0.065)
Treatment *x* IT		-0.050		-0.069
		(0.046)		(0.064)
Treatment *x* Engineering		0.061		0.121**
		(0.044)		(0.061)
Treatment *x* Logistics		-0.041		0.030
		(0.051)		(0.067)
Résumé skills dummy	0.027	0.025	0.027	0.021
	(0.017)	(0.018)	(0.022)	(0.023)
Job functions dummies	Yes	Yes	Yes	Yes
Observations	1,185	1,185	1,185	1,185

Note: Results from a linear regression with the correspondence study data. The dependent variables are dummies equal to one for strict callback in columns (1) and (2) and for broad callback in columns (3) and (4), and zero otherwise. The independent variables are dummies for job functions with 'Finance and accounting' in the reference category. 'Résumé skills dummy' is a variable equal to 1 if the job ad mentions any of the skills highlighted by the candidate in the 'Skills' section of the résumé, and zero otherwise. The names of job functions are shortened for clarity but follow the same order as Table 1. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3 – Percentage of participants correctly guessing which résumé contained AI skills (Prolific study)

	Percentage that guessed correctly	P-value	N
Finance and accounting	97.67%	0.000***	129
HR	94.66%	0.000***	206
Marketing	96.23%	0.000***	53
IT	99.24%	0.000***	132
Engineering	93.80%	0.000***	129
Logistics and supply chains	91.80%	0.000***	122

Note: Results from Prolific study data. Participants are shown two résumés, one with AI skills and one without. The résumés are otherwise identical. They are then asked to recall which résumé contained AI skills (they cannot go back and view the résumés again). They are incentivized with a bonus payment of £0.10 if they guess correctly. Reported are the percentage of participants correctly guessing which of the two résumés contained AI skills, and the *p*-values for a two-sided Binomial test against a random guess of 50%. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4 – Perceived importance of AI skills for execution of the job role (Prolific survey)

	Agree	Disagree	Neither / nor	P-value	Z-score
Finance and accounting	54.26%	22.48%	23.26%	0.000***	8.647
HR	51.94%	22.33%	25.73%	0.000***	10.205
Marketing	56.60%	18.87%	24.53%	0.000***	7.020
IT	52.27%	25.00%	22.73%	0.000***	7.236
Engineering	64.34%	13.18%	22.48%	0.000***	17.178
Logistics and supply chains	66.39%	11.48%	22.13%	0.000***	19.026
All functions	57.07%	19.33%	23.61%	0.000***	26.537

Note: Results from Prolific study data. Results from the question "The AI skills [*list of AI skills from the résumé*] would significantly improve a candidate's ability to execute the [*job function*] assistant role" on a scale from 1, 'Strongly disagree' to 5 'Strongly agree'. 'Agree' and 'Strongly agree' are reported together as 'Agree'; 'Disagree' and 'Strongly disagree' are reported together as 'Disagree'; 'Neither / nor' stands for 'Neither agree nor disagree'. P-values and test statistics (z-scores) from a two-sided one sample test of proportions that tests the difference in the share of the replies between the categories 'Agree' and 'Disagree'. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5 – Perceived under- or over-qualification of candidates by résumé (Prolific survey)

	AI résumé	Control résumé	P-value	Test statistic
Finance and accounting	3.853	3.783	0.319	9
HR	3.791	3.748	0.491	9
Marketing	3.811	3.717	0.557	5
IT	4.030	3.924	0.130	14
Engineering	3.605	3.690	0.228	-11
Logistics and supply chains	3.951	3.861	0.270	11
All functions	3.838	3.790	0.096*	37

Note: Results from Prolific study data. Each row presents the differences in mean response from 1 'Underqualified to 5 'Overqualified' to the question "In your opinion, how well do the candidate's qualification and experience fit an entry-level [*job function*] role?". P-values and test statistics from a two-sided Fisher-Pitman permutation test for paired replicates with 200,000 runs. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

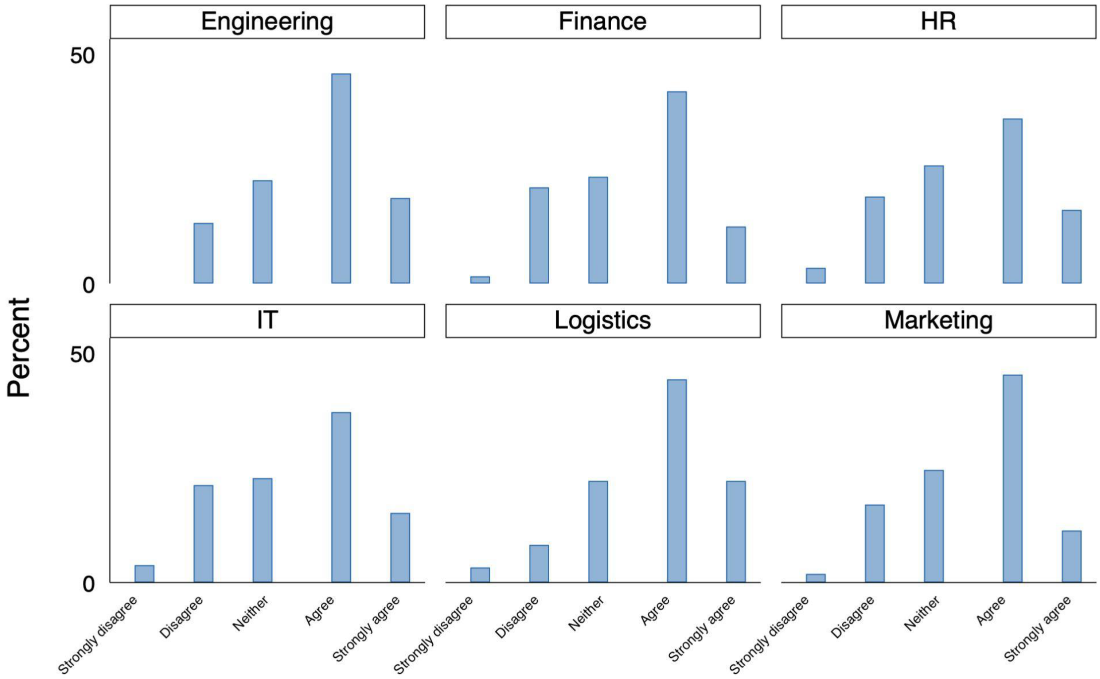
Table A6 – Mean expected strict callbacks by function (Prolific survey)

	AI résumé	Control résumé	P-value	Test statistic
Finance	33.34 (12.52)	30.20 (12.25)	0.000***	405
HR	31.62 (13.02)	29.00 (12.67)	0.001***	539
Marketing	27.68 (14.71)	24.11 (12.75)	0.034**	189
IT	29.70 (13.73)	28.10 (13.42)	0.073*	211
Engineering	32.50 (13.44)	28.51 (13.45)	0.000***	515
Logistics	31.56 (13.00)	29.39 (13.44)	0.005***	265
All functions	31.44	28.69	0.000***	2124

Note: Results from Prolific study data. Participants' response to the question: "Out of 50 companies, how many companies invited the candidate with this résumé for an interview?". We report expected strict callback by function with standard errors in the parentheses. P-values and test statistics from two-sided Fisher-Pitman permutation test for paired replicates. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure A1 – AI skills' relevance for the job (Prolific Survey)

Respondents' answers *to the question "The AI skills [in the résumé] would significantly improve a candidate's ability to execute the [job function] assistant role"*



Note: Results from Prolific survey data. Respondents answer after seeing both AI and control résumés. 'Neither' is short for 'Neither agree nor disagree'.