Humans, Generative AI, and Disparities in Access to Employment*

Silvia Barbareschi
UC Berkeley

Angelica Bozzi
Imperial College London

Francesca Miserocchi
Tilburg University

(JOB MARKET PAPER)

November 10, 2026

Click here for the latest version.

Abstract

How does delegating applicant screening to generative AI change who gets hired? We study an Italian job platform that screens applicants on behalf of thousands of employers and, in 2025, moved screening from human reviewers to an LLM. We observe which candidates progress through each stage of the hiring process, from applications to shortlisting through a first round of screening, and finally the hiring decisions by the employer. We establish three results. First, human screeners discriminate: within the same vacancy, applicants with immigrant-sounding names are 27% less likely to be shortlisted than applicants who are otherwise identical on everything the screener observes on the platform. The penalty differs sharply and persistently across reviewers applying the same rubric, even for the same employer. Second, we compare human shortlisting decisions with AI evaluations of the same applications, holding constant the number of candidates. AI cuts the penalty by 63% and shrinks its dispersion across reviewers' applicant pools. Third, exploiting as good as random variation in the assignment of vacancies to different screening regimes, we provide causal evidence that shifting screening authority to AI, relative to reviewers who already see AI scores, closes the ethnic hiring gap and leads to faster shortlisting without reducing total hires or employer satisfaction. By contrast, the age penalty, which AI also removes at screening, reappears in employers' hiring decisions. AI screening can remove discrimination introduced by intermediaries, but not discrimination rooted in employers' own preferences.

*We are grateful to Sam Asher, Matilde Bombardini, Nina Buchmann, Sydnee Caldwell, Michela Carlana, Viola Corradini, Ernesto Dal Bò, Stefano DellaVigna, Frederico Finan, Stefano Fiorin, Morgan Foy, Roberto Hsu, Kelsey Jack, Samuel Kapon, Patrick Kline, Eliana La Ferrara, Thomas Le Barbanchon, Massimo Morelli, Daniela Paz, Gautam Rao, Simon Quinn, Nina Roussille, Sara Spaziani, Dmitry Taubinsky, Francesco Trebbi, Reed Walker, Chris Walters, as well as participants in the Oliver E. Williamson Seminar, Labor Lunch, and Behavioral Lunch at UC Berkeley, AI & Society Conference at Bocconi University, All California Labor Conference, Advances with Field Experiments (AFE) Conference, Internal Seminar at Imperial College London, Applied Lunch at Bocconi University, Data Science and Simulation Seminar at Luxembourg Institute of Socio-Economic Research (IZA-LISER), 2nd Transatlantic Political Economy Conference at King's College London and at the 2026 OpenAI AGI Economy conference, for helpful comments and conversations. We thank Francesco Mastidoro and Murat Demiraslan for outstanding research assistance. This project was approved by the UC Berkeley Institutional Review Board (IRB 2025-08-18853) and by the Imperial College London Research Ethics Committee (8399897). We gratefully acknowledge funding from the Laboratory for Effective Anti-poverty Policies (LEAP) INSPIRE at Bocconi University and the European Union (MSCA, FAIRWORK).

1 Introduction

Much of the debate about artificial intelligence (AI) and labor markets concerns which jobs and tasks will be replaced by it.[1] However, the effects of AI on employment can arise not only through changes in labor demand, but also through changes in how workers gain access to jobs. Generative AI can now perform tasks traditionally performed by human recruiters: reading résumés, interpreting work histories, and assessing applicants' fit for a position, and adoption is already widespread.[2] These technologies enter a hiring process in which unequal treatment of minority applicants by and across human decision-makers is well documented (Bertrand and Mullainathan, 2004; Neumark, 2018; Quillian et al., 2017; Kline et al., 2022). Because screening determines which applicants advance to employers' consideration, replacing human screening with AI could reshape access to employment, even in occupations whose tasks are not necessarily affected by it. How does delegating screening to AI reshape access to jobs and disparities in the hiring process?

The effects of AI screening on hiring are ambiguous for two reasons. First, human reviewers and AI may select different candidates even when instructed to apply the same criteria. Human reviewers may acquire private information about employers' needs beyond stated job requirements, but their discretion may also reflect personal biases, mistaken expectations of employer preferences, or limited attention (Hoffman et al., 2018). An AI screening rule may reduce these distortions while losing useful contextual information or introducing biases of its own. Second, screening is only one stage of hiring: employers retain the final decision. If screening penalties reflect employers' preferences, employers may reproduce them when choosing whom to hire. If instead they reflect reviewers' own biases or mistaken expectations, changing the screening rule may allow previously excluded candidates to be hired. The consequences of replacing human screening with AI therefore depend on two factors that are rarely observed together: whom each screener selects and how employers respond to the resulting candidate pools.

To study this question, we partner with an Italian job-matching platform that acts as an intermediary between job seekers and employers in the hospitality sector, and, in 2025, delegated screening to a large language model (LLM). The platform screens applicants on behalf of thousands of employers and records their subsequent decisions. This allows us to

[1] See Acemoglu et al. (2022); Hampole et al. (2025); Klein Teeselink (2025); Brynjolfsson et al. (2025a); Eloundou et al. (2024).

[2] LinkedIn's *Future of Recruiting 2025*, based on a survey of more than 1,000 talent professionals, reports that 37 percent of respondents' organizations were integrating or experimenting with generative AI in recruiting, up from 27 percent a year earlier.

observe both whom the intermediary selects and whom employers ultimately hire. About 45 percent of applicants have immigrant-sounding names, which we use as a proxy for perceived immigrant background.[3] This potentially makes ethnic disparities an economically important dimension along which to study how AI recruitment changes access to jobs.[4] Human reviewers interact with employers and assist with vacancy posting, creating scope to acquire information about their needs beyond recorded job requirements. Applicants submit a CV, state their job preferences, and complete a standardized WhatsApp assessment.[5] The platform then decides whom to forward, and employers can interview and hire only among forwarded applicants. Human reviewers and the LLM screener apply the same written rubric to the same applicant information; the LLM is a general-purpose model instructed through a prompt and not trained on the platform's past decisions.[6] This allows us to study whether the same stated criteria generate different selections when applied by human reviewers and an LLM. Crucially, the platform records the entire recruitment pipeline.

Our data allow us to reconstruct the employers' choice sets and follow how employers select from them, overcoming data limitations from previous research. The administrative records from the platform cover approximately 320,000 applications to 4,949 vacancies posted by 1,433 employers. They combine detailed applicant characteristics and WhatsApp assessments with vacancy requirements and characteristics, measures of applicant–job fit, and employer characteristics. Each application is linked to its screening decision, the human reviewer and/or AI screener responsible for that decision, and employers' subsequent interview and hiring decisions. We also observe the positions of the candidates on the shortlist and whether employers open their profiles. We complement these records with employer satisfaction and measures of match stability for subsets of the sample. The records cover both human screening before AI adoption and the subsequent deployment.

We combine three research designs. First, we measure where disparities arise under human screening by following applicants to the same vacancy from shortlisting to hiring, holding fixed detailed applicant characteristics, CV presentation, applicant–job fit, WhatsApp as-

[3]We classify names using NamSor. The indicator does not measure citizenship, migration history, or work authorization. Section 2.3 describes the measure and its validation.

[4]Research studies in Italy document unequal treatment of applicants of immigrant origin in recruitment (Allasino et al., 2004; Busetta et al., 2018).

[5]A chatbot runs the assessment right after the application and asks a fixed set of questions mostly to check Italian language skills, availability, and motivation. It does not make shortlisting decisions.

[6]This implementation potentially reduces the data requirements and development costs of adopting algorithmic screening. This makes the approach relevant to organizations beyond those with the resources to build their own hiring models.

sessments (which includes Italian language skills), and application timing.[7] Second, we compare the two screening rules on the same applicants at the same capacity. We run the LLM screener ex post on applications previously evaluated by human reviewers. Within each vacancy, we construct a counterfactual shortlist by selecting the highest-scoring applicants under AI, holding the applicant pool and shortlist size fixed. Third, we exploit the platform's as good as random assignment of job postings to screening regimes during the transition to AI.

Employers were not told which regime screened their postings. Our main comparison exogenously moves screening authority from reviewers who see AI scores but decide themselves to a hybrid regime in which the AI automatically forwards high-scoring applicants and rejects low-scoring ones, while reviewers decide intermediate cases. This comparison captures the major shift in decision authority, and exogenously changes employers' choice set.

We find that, under human screening, ethnic disparities emerge at shortlisting, persist into hiring, and vary substantially across reviewers and employers. Applicants with immigrant-sounding names are 27.4 percent less likely to be shortlisted and 36.5 percent less likely to be hired than observationally comparable applicants with Italian-sounding names applying to the same vacancy. Conditional on shortlisting, interview and hiring gaps are small and imprecisely estimated, consistent with the screening stage—which determines whether a candidate enters the employer's choice set—playing an important role in generating disparities. These average disparities hide substantial heterogeneity across reviewers and employers, which persist when we estimate them jointly and control for detailed applicant characteristics and applicant–vacancy fit. Controlling for reviewer workload and screening experience does little to reduce the dispersion across reviewers. Human screening also disadvantages older applicants, providing a complementary demographic comparison.

Comparing AI and humans at screening, we find that AI reduces the ethnic shortlisting penalty conditional on observables by 63 percent. AI also eliminates the older age disadvantage. AI and humans disagree on whom to screen in: among applicants shortlisted by at least one rule, only 33.1 percent are selected by both. Across the applicant pools originally assigned to different human reviewers, estimated ethnic shortlisting penalties are less dispersed under AI. These changes accompany broader differences in the characteristics

[7]The *conditional gap* compares applicants to the same vacancy, holding fixed the information available to the screener. Under the assumption of selection on observables, it measures direct discrimination (Bohren et al., 2025), the object correspondence studies identify experimentally. This assumption is particularly plausible in our setting because reviewers assess applicants through the platform, do not have additional interaction with them, and our data record the applicant information available to them, including CVs, stated preferences, and WhatsApp assessments.

associated with selection. Flexible predictive models assign greater predictive importance to hospitality experience under AI, whereas age and immigrant-sounding names are more important predictors of human decisions. Thus, even when instructed to apply the same criteria and selecting the same number of candidates, humans and AI construct substantially different employer choice sets, with different ethnic and age-related penalties. Whether these changes translate into different hiring outcomes depends on how employers respond to the candidates they receive.

Partial delegation of screening authority to AI substantially and significantly narrows the ethnic hiring gap: the conditional hiring gap moves from -0.328 to $+0.052$ percentage points, a change of 0.380 points ($p = 0.046$). Moreover, the time to shortlisting falls by 1.33 days, the probability that a posting generates a hire does not change detectably, and we do not find any deterioration in measures of employer satisfaction, match quality, or available measures of separation rates. These findings are consistent with AI producing some efficiency gains in screening, and with human reviewers imposing personal penalties or incorrectly anticipating employer preferences. Age differs: disadvantages attenuated at screening by AI re-emerge in employers' choices. Replacing human screening with a rule that generates smaller disparities can thus remove barriers introduced by intermediaries, while others re-emerge later in hiring. Additional comparisons among screening regimes show no detectable reduction in ethnic or age hiring disparities when human reviewers receive AI scores but retain screening authority, and no detectable change in the main hiring outcomes when AI screening is extended to intermediate cases previously subject to human review. These comparisons highlight the distinction between providing algorithmic information and delegating decision authority: the consequences of AI depend on the role it is given in the hiring process.

The mechanism evidence suggests that changes in applicant characteristics and job fit contribute to the narrowing of the ethnic hiring gap, while improved relative ranking and larger shortlists alone do not appear to drive the result. We find no detectable improvement in the relative ranking or profile-opening probability of applicants with immigrant-sounding names, and the estimated hiring-gap reduction remains largely unchanged when controlling for rank and AI scores or reweighting postings on shortlist size. If we compare the human shortlists with the top AI selected applicants (matched to the size of the human shortlists) we find that the AI selects applicants with more hospitality experience in both name groups, and there is no difference in the an index of applicant-job fit. Applicants with immigrant-sounding names also have fewer employment gaps, and their measure of motivation is higher, suggesting an improvement in quality in both technical and soft skills for this group. Shortlist expansion

may serve a different role: employers hire candidates beyond the matched-size set in both groups, suggesting that the AI ranking does not fully capture employers' needs, and that broader choice sets might provide additional matching opportunities and may help sustain overall hiring. Also, while hiring from the expansion set does not seem to drive the reduction in the ethnic hiring gap, our decomposition suggests it sustains the age penalty. These observed hires do not establish whether employers would otherwise have substituted toward matched-size set candidates or left vacancies unfilled. However, they suggest a complementarity between algorithmic screening and human judgment at the final selection stage. This might help to sustain overall hiring but also to sustain disparities attenuated at screening.

Broadly, our findings show that disparities generated upstream can affect disparities downstream, that AI screening can reshape access to employment, and that it is a change in the allocation of opportunities, not merely a reduction in the cost of processing applications.[8]Our findings also suggest a need to rethink how disparities arise and operate when we move from heterogeneous human judgments to AI screening applying a common screening rule. A common rule can reduce disparities arising from individual discretion while making any remaining penalties more systematic across employers. Inequality may also take a different form through the criteria used to rank candidates and the choice sets presented to employers. Its consequences therefore depend not only on how the algorithm screens, but also on whom employers consider and ultimately hire.

First, we contribute to research on AI and human decision-making. Prior work compares human judgment with predictive algorithms, highlighting the trade-off between consistent decision rules and the use of decision-makers' private information (Kleinberg et al., 2018; Angelova et al., 2026). In hiring, Hoffman et al. (2018) find that greater managerial discretion in departing from job-test recommendations is associated with shorter worker tenure, while Li et al. (2026) show how incorporating exploration into predictive screening algorithms can improve demographic diversity and hiring potential. Recent research examines generative AI assistance in professional settings (Brynjolfsson et al., 2025b; Mehmood et al., 2026), disparities in AI-generated evaluations (Leibbrandt et al., 2026), and AI delegation in teacher recruitment (Awuah et al., 2026). To the best of our knowledge, we are the first to examine how a real-world transition to LLM screening changes demographic disparities throughout a multi-stage hiring process, including actual hiring outcomes when employers retain the final decision.

Second, we contribute to research on discrimination in multistage systems. Disparities

[8]AI evaluates candidates at a much lower marginal cost (roughly =0.01c per evaluation versus about =1.50c for human review).

arising at one stage can be perpetuated, mitigated, or amplified by subsequent decision-makers (Bohren et al., 2025; Baron et al., 2024; Harrington and Shaffer, 2026). This literature examines how disparities arise and propagate across stages; we contribute by showing how changing an upstream screening rule affects hiring when employers retain the final decision. The new rule reduces screening penalties for both applicants with immigrant-sounding names and older applicants, but their downstream outcomes diverge: the ethnic hiring gap narrows, while age-related disadvantages persist in employers' choices. These findings suggest that the effects of upstream reforms depend on whether screening penalties reflect intermediaries' own judgments or anticipate employers' preferences.

Our analysis also connects to research on employer search, recruitment intermediaries, and attention allocation (Horton, 2017; Fernando et al., 2023; Bartoš et al., 2016). We examine how employers respond when a change in screening technology alters the candidates available for consideration. Evidence on shortlist size, applicant characteristics and job fit, and ranking helps characterize how broader access translates into employer consideration and hiring, while employer satisfaction and separation outcomes provide complementary evidence on whether improved access comes at a cost to employers or match stability. This connects the distributional consequences of screening to employers' use of the resulting choice sets and the outcomes of the resulting matches.

From a policy perspective, these findings complement research on policies aimed at studying and reducing hiring disparities (Mocanu, 2024). Goldin and Rouse (2000) find that blind auditions improve women's advancement and hiring prospects. More recently, Behaghel et al. (2015) find that anonymizing résumés actually reduces minority candidates' interview and hiring probabilities among participating firms. Our contribution is to study a different procedure that leaves applicant identity visible throughout the process. Delegating screening to an AI instructed to use the same criteria as humans, changes the characteristics of the applicant pool visibile to employers. Our results show that partial delegation to AI can narrow ethnic hiring disparities even when employers retain both identity information and authority over final selection. In addition our findings inform the more recent design of human oversight in algorithmic hiring.[9] Human discretion may introduce group penalties, but it may also preserve information about employer needs that a common screening rule misses. Effective oversight therefore requires specifying which decisions humans retain and evaluating their consequences for both access to employment and match outcomes. As common screen-

[9]The European Union's AI Act includes recruitment and candidate-selection systems within its high-risk framework and establishes requirements for transparency and human oversight; see Regulation (EU) 2024/1689, Article 6, Annex III, point 4(a), and Articles 13–14.

ing rules spread across employers, this evaluation should also consider whether penalties embedded in those rules expose applicants to repeated disadvantages across vacancies.

The paper proceeds as follows. Section 2 describes the institutional setting, platform, and data. Section 3 outlines a conceptual framework. Section 4 studies disparities in the human screening. Section 5 compares human and AI screening on the same applications at equal shortlist capacity. Sections 6 and 7 examine deployment outcomes and the mechanisms behind our results. Section 8 examines the other comparisons between screening regimes. Section 9 discusses the findings. Section 10 concludes.

.

2 Institutional Setting, AI Technology, and Data

2.1 Institutional Setting: the Recruitment Process

We collaborate with a major online job-matching platform operating in the Italian hospitality sector. Hospitality is an important source of employment in Italy, with non-Italian workers accounting for approximately 15 percent of employment in the sector (ISTAT, 2022).[10] Therefore, the hospitality sector provides a salient setting in which to study ethnic disparities in access to employment. The platform acts as an intermediary between workers and employers: it screens applicants on behalf of employers, who retain control over interviews and hiring. Intermediaries play an important role in the Italian labor market. In 2025, 11.4 percent of firms that had hired in the previous year reported using employment agencies or recruitment firms (Unioncamere and Ministero del Lavoro e delle Politiche Sociali, 2025).[11]

The platform operates as a subscription-based marketplace in which employers pay a fixed monthly fee of approximately EUR 499 to access prescreened candidates. During our study period, it had more than 200,000 registered job seekers and over 1,400 employers, with approximately 10,000 active job seekers per month. At its peak, around 300 employers posted approximately 600 vacancies per month. Vacancies span 79 occupations, from head chefs to dishwashers, with substantial variation in compensation both across and within occupations.[12] The platform controls which applicants are forwarded to employers and keeps

[10] In 2022, Italy's accommodation and food-service sector comprised approximately 329,000 enterprises and employed 1.56 million people (ISTAT, 2025b).

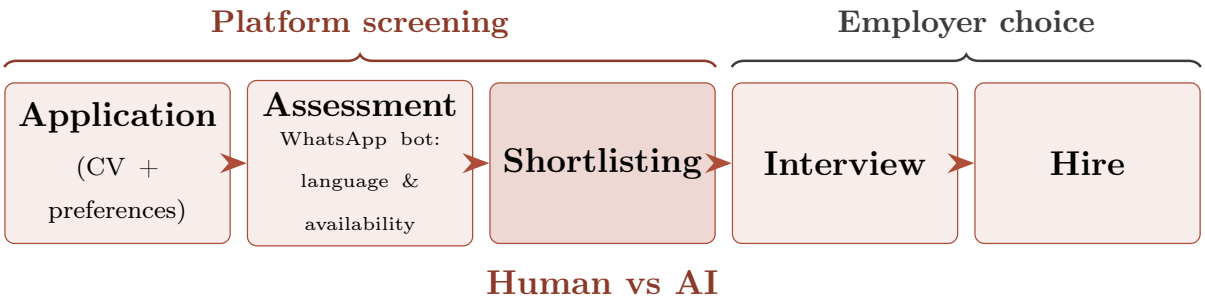
[11] 2025 Excelsior survey. Firms could check multiple recruitment channels. Online screening platforms are not separated from other types of intermediaries.

[12] Among full-time positions, 44 percent of the variation in average annual salaries is between occupations and 56 percent within occupations (Appendix E).

track of the full recruitment pipeline, from applications to screening decisions and subsequent employer choices.

Figure 1 illustrates the recruitment process and the division of authority. To apply to a specific vacancy, job seekers submit a CV and state their preferences over sectors, working hours, salary, and shifts. Since November 2024, for each application submitted, job seekers need to complete a standardized WhatsApp assessment that is meant to collect additional information on availability, willingness to relocate or commute, motivation, and Italian proficiency.[13]Till May 2025, all the applications were evaluated by the platform's human reviewers who decided which applications send to employers. Employers then could act only on this forwarded set: they opened candidate profiles, selected whom to interview, and recorded the eventual hire. Reviewers were full-time employees and vacancies were assigned to them primarily according to their availability. Returning employers were more likely to work with the same reviewer. These interactions could provide reviewers with employer-specific information beyond the requirements recorded in job postings. Importantly, the reviewer shortlisting decisions determined which applicants entered the employer's choice set in the first place.

Figure 1: Recruitment pipeline.



Recruitment in Italy is subject to legal restrictions on discrimination. Italian law prohibits discrimination based on racial or ethnic origin, sex, and age, subject to specific exceptions.[14] On the platform, employers cannot restrict vacancies by age, gender, or nationality. Nevertheless, experimental studies in Italy document unequal treatment of applicants of immigrant origin in recruitment (Allasino et al., 2004; Busetta et al., 2018).

[13]Right after applying, candidates receive a WhatsApp message from a chatbot that asks hat asks them a few questions. These responses are converted into structured measures available to screeners.The chatbot does not make shortlisting decisions.

[14]See Legislative Decree 215/2003, Articles 2–3; Legislative Decree 216/2003; and Legislative Decree 198/2006, Article 27. Legislative Decree 276/2003, Article 10, also prohibits discriminatory preselection by authorized or accredited employment intermediaries. These protections are distinct from rules governing citizenship and authorization to work.

Importantly, immigrant origin is distinct from citizenship and eligibility to work. First, Italian citizenship is not required for employment. Second, EU citizens can work in Italy even without a work permit, while non-EU citizens generally require residence status authorizing employment.[15] Moreover, our main measure of perceived immigrant background is based on whether an applicant has an immigrant-sounding name, as described in Section 2.3. It does not establish citizenship, migration history, or authorization to work.[16]

2.2 The AI screening tool

During 2025, the platform gradually introduced AI screening (Figure 2, and Section 6).

The AI screener uses a general-purpose large language model (GPT-4.1) to evaluate how well an application fits a vacancy. The platform developed a prompt that specifies the evaluation rubric and instructs the model how to apply it. Human reviewers were instructed to follow the same rubric. The agent receives as inputs the vacancy description, the applicant's CV, distance to the workplace, and the standardized WhatsApp assessment transcript, and finally it is asked to produce as output a continuous numerical score between 1 and 10, with higher scores indicating better assessed fit.

The final score is a weighted average of other three sub-scores, created from a total of nine human written criteria. *Job Match* receives a weight of 40 percent and includes experience in the role, required language proficiency, and availability. *Soft Skills and Attitude* receives 30 percent and includes professionalism, motivation, and reliability. *Technical and Logistical Fit* receives the remaining 30 percent and includes practical skills, job stability, and distance from the workplace. Each criterion is scored from 1 to 10 according to explicit guidelines. According to predetermined thresholds of the final score decided by the platform, applications are categorized in three categories: "Send," "Wait," and "Reject." Appendix D reports the nine criteria, their weights, and illustrative scoring anchors.

Unlike predictive algorithms trained on historical hiring decisions (Kleinberg et al., 2018; Li et al., 2026), the model is not trained on the platform's previous applications, screening decisions, or hiring outcomes.[17] Instead, the platform adapts a general-purpose language model through instructions and a fixed rubric. Although reviewers and AI are instructed to

[15] Several types of permits authorize employment, including those for work or family reasons (European Commission).

[16] Only 1.16 percent of CVs with available text explicitly mention a residence or work permit. The lack of this information does not imply that an applicant lacks authorization to work

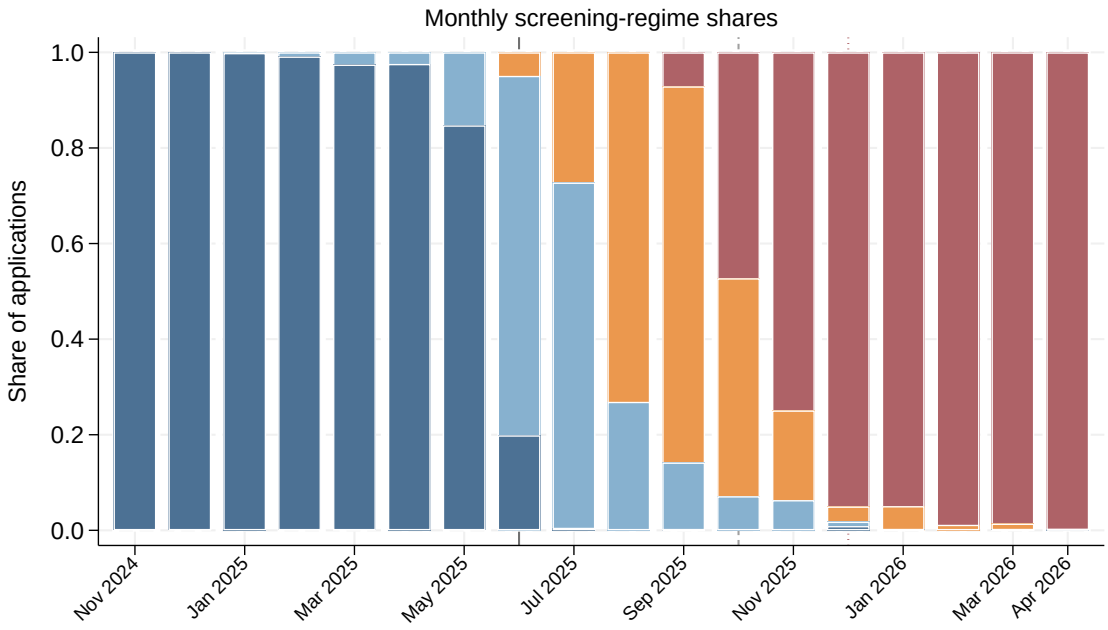
[17] Since the model is not trained on the platform's historical data, past screening or hiring decisions cannot explain differences across applicant groups. However, they could still exist and come from the training data of the general purpose language model. This is important for policy.

apply the same criteria, they may interpret and weigh applicant/applications information differently. In addition, differently from a common algorithmic rule, human reviewers may also use employer-specific information acquired through previous interactions. Such discretion can add useful contextual information (Dal Bó et al., 2021; Voth and Xu, 2022), but it may also reflect reviewers' own preferences or mistaken expectations about employers. Delegating screening to AI may therefore reduce discretionary disparities while changing how employer-specific information enters selection.

This implementation is relevant beyond the platform we study. Unlike proprietary prediction systems that require historical training data and substantial development resources, general-purpose language models can be adapted to screening tasks through prompts and evaluation rubrics. Our setting therefore illustrates a form of algorithmic hiring that may be accessible to organizations without the resources to develop their own prediction models. A unique feature of our setting is that we know exactly the prompt, the model, and the information sent.

.

Figure 2: The transition to AI screening.



This figure shows the transition from human screening to AI screening. On the X axis the month-year, on the Y axis the share of applications assigned to each regime. Section 6 defines the periods and comparison samples used in the analysis.

Human only: human reviewers retain full authority over whom to shortlist. They do not receive any additional information.

Human + Information: human reviewers retain full authority over whom to shortlist. They receive as additional information the decision of the AI screener.

AI-Hybrid: the AI screener automatically advances high-scoring applications, rejects low-scoring applications, while human reviewers retain a role in the intermediate range and can intervene.

Fully AI: the AI screener has full authority in shortlisting.[a]

[a] A designated human remains responsible for overseeing the system.

2.3 Data

Our analysis uses detailed administrative records from the platform covering more than 360,000 applications submitted since November 1, 2024, when the platform introduced the standardized WhatsApp assessment. The data link applicants to job postings, employers, human reviewers, screening decisions, and downstream outcomes, including profile views, interviews, hires, and some post hiring outcomes such as employer satisfaction and separation rates. Table B1 summarizes the main data sources and their coverage.

Worker characteristics. For each worker, we observe demographic characteristics (age, gender, and location), employment status, transport availability, and stated job preferences collected at registration and during the application process. Uploaded CVs provide detailed measures of employment history, education, and language proficiency. We also construct measures of CV quality and presentation, including the presence of a photograph, document structure, length, use of the CV standardized templates such as Europass, and language consistency, as well as the organization of job-relevant information, such as whether the required occupation appears on the first page. Appendix B describes the construction of these variables.

For each application we also observe text conversations from the WhatsApp pre-assessment. We conducted sentiment analysis using Gemma 4, an open-weight language model, to construct measures of availability, shift availability, Italian and English proficiency, relevant experience, distance, soft skills, technical skills, motivation, and response quality. These measures follow the dimensions of the AI screening rubric but are constructed independently of the production AI score. Appendix C describes the procedure and variable construction. We also know the reviewer that evaluated each application, the time in which each application was submitted and the order in which reviewers evaluated them.

Our main measure of perceived immigrant background is based on the country of origin inferred from applicants' names using NamSor (Bursztyn et al., 2024). We classify a name as Italian-sounding if either NamSor's primary or alternative country prediction is Italy, and as immigrant-sounding otherwise. This measure captures the background conveyed by a name rather than citizenship or migration history. Among 55,116 applicants with both a name classification and CV-stated nationality, the two measures agree in 86.4 percent of cases.[18] Appendix B.1 provides validation details, while Appendix F reports robustness checks using an alternative measure incorporating CV-stated nationality.

Posting and employer characteristics. For each vacancy, we observe occupation (79 detailed categories), compensation (salary and benefits), contract terms (contract type, full-time status, and shifts), location, and openness to non-local candidates. We also observe employer characteristics, including firm size. We link approximately 58 percent of employers to AIDA firm records (Moody's Analytics, formerly Bureau van Dijk), which provide information on employment, revenues, assets, profits, labor costs, and firm management. As

[18]Sensitivity for CV recorded non-Italians is 98.0 percent and specificity for CV recorded Italians is 79.1 percent. Differences may reflect naturalization, dual nationality, second generation status, or classification error. CV nationality is itself an imperfect benchmark for the origin perceived from a name.

these data are available only through 2024, before AI deployment, we use them as baseline covariates but cannot examine firm-level outcomes following AI adoption.

Combining worker and posting characteristics, we construct measures of applicant–vacancy compatibility, including occupation, experience, language, distance, and sector match. Their average forms a match index (referred to as *matched composite* index throughout the analysis) ranging from zero to one. Appendix B.1 provides additional details on the construction of these variables.

Outcomes. For each application, we observe whether the applicant is shortlisted, whether the employer opens the applicant's profile or indicates interest, and whether the applicant is interviewed or hired. Under algorithmic screening, we also observe the AI score. At the posting level, we measure shortlist size, screening time, and whether the employer interviews or hires at least one applicant.

- Retention and employer satisfaction: we additionally observe employer satisfaction from a survey administered by the platform two to three weeks after a posting closes, regardless of whether a candidate was hired. For a limited sample of hires, a follow-up survey records whether the employment relationship has ended within specified horizons. The response rate for this survey is approximately 30 percent. Because the job satisfaction and retention measures are incomplete and potentially subject to selection, we use them as supplementary evidence, and we use also different measures for separation: administrative records identify worker reapplication for another position and employer reposting of the same role (Appendix I).

Descriptive statistics. Appendix Tables E1, E2, and E3 compare applicant characteristics, applicant–vacancy match, and vacancy sorting between applicants with Italian- and immigrant-sounding names. Applicants with immigrant-sounding names are younger (30 versus 35 years on average) and have fewer years of recorded hospitality experience (3 versus 6 years). They are less likely to report employment gaps (47 versus 62 percent) and more likely to report high English proficiency (61 versus 55 percent), but less likely to report high Italian proficiency (69 versus 100 percent). They also apply more frequently to lower-skilled and lower-paid positions, are less likely to apply to permanent jobs, and apply to vacancies requiring less prior experience. Their average observed applicant–vacancy match is lower, particularly for occupation and minimum-experience requirements. These differences motivate our detailed applicant and match controls and comparisons within job postings.

Demographic predictors of shortlisting. Age and immigrant sounding names are among the strongest demographic predictors of shortlisting under human screening (Appendix Figure F5). We assess predictive importance using XGBoost, which allows nonlinear relationships and interactions among observed characteristics. Then SHAP values attribute each prediction's deviation from a reference value to the model's inputs, giving the relative importance of each input. Figure F5) ranks variables by their average absolute contribution to predictions for applications from postings excluded from model training. These contributions measure predictive importance, not causal effects, and do not indicate whether a characteristic raises or lowers shortlisting probability. However, this evidence motivates our focus on ethnic disparities, with age providing an additional demographic comparison.

3 Conceptual Framework

Hiring on the platform has two stages. A screener, human or AI, forms a shortlist from the applicant pool; the employer then chooses whom to interview and hire, or leaves the vacancy unfilled. We assume that the platform and employers share the objective of successful matches. Screening and hiring may diverge because screeners assess applicants with imperfect information and may apply group adjustments that differ from employers' own (Hoffman et al., 2018; Harrington and Shaffer, 2026).

Under screening regime $r \in \{H, \text{AI}\}$, the screener's assessment of applicant i for vacancy j is

$$s_{ij}^r = x_{ij}'\omega_r + \lambda_r u_{ij} + b_{rj} G_i + \varepsilon_{ij}^r,$$

where x_{ij} contains recorded applicant–vacancy characteristics, u_{ij} captures employer-specific fit not recorded in the vacancy or observed in our data, G_i indicates group membership, and ε_{ij}^r is assessment noise. The screener forwards the K_j^r highest-scoring applicants. Human reviewers may learn about employers' needs through repeated interactions, allowing them to incorporate information about u_{ij} (potentially higher λ_H). Their group adjustment b_{Hj} can vary across reviewers and employers, reflecting beliefs about employers' preferences, reviewers' own preferences, or error. AI applies a common rubric, with a group adjustment b_{AI} that is constant across vacancies and potentially less information about employer-specific fit (lower λ_{AI}). Standardization can therefore reduce variation in group-specific adjustments while discarding useful information. Because the two rules may also weigh recorded characteristics differently ($\omega_H \neq \omega_{\text{AI}}$), they can select different applicants at equal capacity without either necessarily producing better matches.

Employers determine which screening changes persist. An employer hires the shortlisted candidate with the highest value $\tilde{\pi}_{e,ij} - t_e G_i$, where $\tilde{\pi}_{e,ij}$ is its assessment of fit and t_e captures a group-specific adjustment, provided this exceeds the value of leaving the vacancy unfilled. Hiring disparities can arise directly from employers' choices among shortlisted candidates or through screening, which determines whom employers consider (Bohren et al., 2025). If human screening excludes candidates whom employers would otherwise hire, changing the screening rule can narrow hiring gaps. If employers themselves disadvantage these candidates, disparities may persist downstream. We examine these patterns for perceived immigrant origin and age.

AI screener deployment may also change the shortlist size. When screener and employer assessments differ, a longer shortlist may include candidates employers prefer to those ranked highest by the screener. With nested lists and costless consideration, expansion cannot worsen the employer's best available option, but it need not increase hiring if employers simply substitute among candidates. With limited attention, larger shortlists may instead change which applicants employers examine (Bartoš et al., 2016).

These mechanisms motivate two complementary analyses. We first compare human and AI selection for the same applicants at a fixed shortlist size, isolating differences in screening rules. We then examine AI deployment, which changes the screening regime, shortlist size, and screening time together, and trace its consequences through employer consideration and hiring. Appendix K presents the formal model.

4 Human Screening

This section analyzes disparities in the human selection recruitment process. Two main findings emerge. First, even after accounting for detailed application characteristics, applicants with immigrant-sounding names face a large shortlisting penalty that persists in hiring, while gaps among shortlisted applicants are small. Second, the penalty varies substantially across reviewers and employers, even after accounting for observed applicant and vacancy characteristics. We observe similar patterns for age disparities, but no meaningful gender differences. We begin by documenting unconditional disparities across the hiring process under human selection (Appendix Figure F1), then investigate possible mechanisms. These results provide the human screening benchmark against which we evaluate algorithmic screening in the later sections.

4.1 Where Do Disparities Emerge?

Applicants with immigrant-sounding names are 48 percent less likely to be shortlisted and approximately 55 percent less likely to be interviewed or hired than those with Italian-sounding names. Among shortlisted applicants, these gaps fall to 13 percent for interviews and 12 percent for hires (Appendix Figure F1). These raw differences may reflect differences in applicant characteristics and sorting across vacancies.

Ethnic disparities in shortlisting We therefore compare applications within the same vacancy, accounting for observed applicant qualifications, candidate-job match, and WhatsApp pre-assessment information:

$$Y_i = \alpha_j + \lambda_t + \beta G_i + X_i'\gamma + M_i'\theta + \varepsilon_i. \tag{1}$$

where Y_i is the binary outcome for screening, interview, or hiring for the application i; j denotes the job posting and t denotes the month-year associated with application i. G_i is the immigrant-sounding-name indicator. X_i contains applicant and CV characteristics; M_i contains applicant–vacancy match measures, WhatsApp assessment scores, and application-order controls. Posting fixed effects α_j compare applicants within the same vacancy, while month-of-application effects $\mu_{t(i)}$ absorb common calendar changes. Standard errors are clustered by posting.

The β coefficient can be interpreted as the effect of perceived origin conveyed by a name, holding constant observed characteristics.[19] A causal interpretation of β requires a selection on observables assumption: conditional on the included controls, name classification must be unrelated to unobserved determinants of the screening decision. Our setting provides unusually rich information on applications. Nevertheless, reviewers may extract information from CVs and transcripts that our constructed measures do not capture.

[19] Because these characteristics may themselves reflect past discrimination, this conditional effect does not capture the total disparity associated with perceived immigrant background (Bohren et al., 2025).

Table 1: The shortlisting gap persists after accounting for observables

	Dep. var.: Shortlisted (0/1)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Immigrant-sounding name	-0.066***	-0.062***	-0.045***	-0.048***	-0.044***	-0.042***	-0.039***	-0.036***	-0.037***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Match composite								0.168***	
								(0.005)	
Job-posting and Month-Year FE	No	✓	✓	✓	✓	✓	✓	✓	✓
WhatsApp assessment scores	No	No	✓	✓	✓	✓	✓	✓	✓
Demographics	No	No	No	✓	✓	✓	✓	✓	✓
Experience	No	No	No	No	✓	✓	✓	✓	✓
CV quality + attention	No	No	No	No	No	✓	✓	✓	✓
Preferences	No	No	No	No	No	No	✓	✓	✓
Match dimensions + order	No	No	No	No	No	No	No	No	✓
Match composite + order	No	No	No	No	No	No	No	✓	No
Applications	123,326	123,311	123,311	123,311	123,311	123,311	123,311	123,311	123,311
Mean Y (Italian-sounding)	0.137	0.137	0.137	0.137	0.137	0.137	0.137	0.137	0.137
Adj. R^2	0.011	0.068	0.115	0.127	0.134	0.137	0.139	0.150	0.158

Notes. This table reports regression estimates of Equation (1). The dependent variable is an indicator for whether the applicant was shortlisted. The reference group are applications with Italian-sounding names. Columns (1)–(9) progressively add controls as shown in the table, Section 2.3 and Appendix B.1 describe their construction. Column (8) replaces the five separate applicant–job match controls with their mean across observed components. Column (9) reports the preferred specification. Missing control values are set to zero and accompanied by missing-information indicators. Standard errors are clustered at the posting level, and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 1 shows how the shortlisting gap changes as we progressively add controls. Without controls, applicants with immigrant-sounding names are 6.6 percentage points less likely to be shortlisted, or approximately 48 percent less likely than applicants with Italian-sounding names (column 1). The gap falls to 6.2 percentage points after adding posting and month–year fixed effects, and to 4.5 points (33 percent) after including WhatsApp assessment scores (columns 2–3). With the full set of controls, the gap remains 3.7 percentage points, or 27 percent of the shortlisting rate for applicants with Italian-sounding names (column 9). Observed characteristics thus account for less than half of the raw disparity.

The penalty also varies by perceived origin. Relative to the shortlisting rate of applicants with Italian-sounding names, the conditional penalty is approximately 63 percent for South Asian names, 47 percent for Sub-Saharan African names, and 43 percent for Arab-country names. It is smaller for Eastern, Central, or Southeastern European names (15 percent), or Latin American or Caribbean names (9 percent) (Appendix Table F5).

The gap is similar under an alternative origin classification that uses CV-stated nationality primarily and then NamSor when nationality is missing (3.6 percentage points; Appendix Table F3). It is also robust to alternative functional forms: specifications with flexible age and WhatsApp controls and cross-fitted XGBoost adjustment yield gaps of 3.8–4.0 points (Appendix Tables F4 and F10).

We next examine potential mechanisms underlying the name penalty.[20] The gap is smaller in customer-facing roles than in other positions (3.1 versus 4.6 percentage points; Appendix Table F6), contrary to the prediction that anticipated customer discrimination generates larger penalties in customer-facing jobs.[21] Relative gaps are larger in Southern Italy and the islands than in the North, potentially revealing differences in local employer preferences (Appendix Table F6). The gap is also larger when CV experience information is missing (5.1 versus 3.7 points; Appendix Tables F13 and F14), consistent with greater reliance on perceived origin when information is scarce.[22] We next assess whether the immigrant gaps can be explained by order and workload effects. We find no systematic relationship between the gap and application order (Appendix Tables F7 and F8). However, immigrants

[20] Appendix F.1 provides the complete analyses.

[21] The smaller gap in customer-facing roles is also confirmed in relative terms using Poisson Pseudo-Maximum Likelihood (PPML), although the difference is no longer statistically significant when vacancy characteristics are interacted jointly. PPML models the conditional shortlisting probability using an exponential mean function, measuring relative rather than percentage-point gaps. A name coefficient β corresponds to a relative gap of $100[\exp(\beta) - 1]$ percent. The estimator can be used with a binary outcome and does not require a Poisson outcome distribution.

[22] The difference remains after adjusting for multiple testing and when interactions with all five missing-information indicators enter jointly. Missing information may also correlate with unobserved applicant characteristics, so this association does not establish that providing more information would reduce the gap.

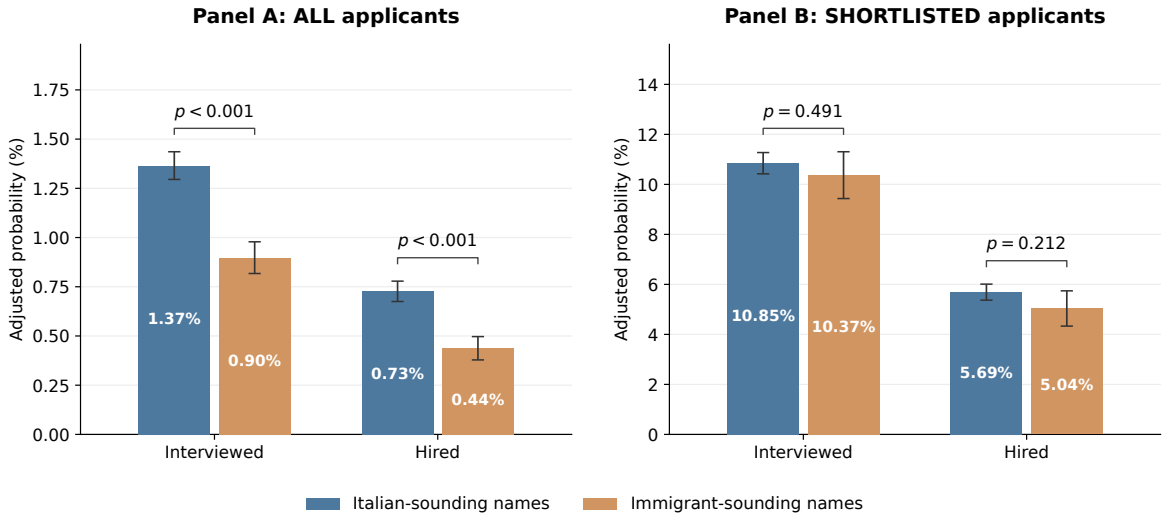
on average apply later, and order is important for shortlisting probability, this is another margin on which AI screening could differ. Higher reviewer workload is associated with a 0.46 percentage point larger gap per standard deviation, but the estimate falls to 0.35 points and becomes imprecise when allowing reviewer-specific baseline name gaps (Appendix Table F9).[23]

Ethnic Disparities in downstream outcomes. Figure 3 applies Equation (1) to interviewing and hiring, using the full applicant pool (Panel A) and using only the shortlisted pool i.e the visible set available to employers (Panel B). Conditional on observables, using the full applicant pool the gaps are negative and statistically significant also in downstream stages: applicants with immigrant-sounding names are 30 percent less likely to be interviewed, and 37 percent less likely to be hired than observationally comparable applicants with Italian-sounding names. Thus, the large shortlisting gap persists throughout the hiring pipeline. Among shortlisted applicants, interview and hiring gaps are smaller in relative terms. Applicants with immigrant-sounding names have approximately 4 percent lower interview probabilities and 12 percent lower hiring probabilities than observationally comparable applicants with Italian-sounding names, and neither difference is statistically significant (Appendix Table F1). These patterns suggest that differential access to employers at the shortlisting stage plays an important role in hiring disparities. However, these absence of statistically significant downstream gaps do not rule out employer penalties. If reviewers penalize applicants with immigrant-sounding names at screening, shortlisted applicants may already be positively selected on characteristics employers value, potentially hiding penalties in subsequent decisions.

We also examine post-hire outcomes to assess differences in match stability across name groups, in the spirit of outcome tests of discrimination (Becker, 1957; Knowles et al., 2001; Arnold et al., 2022a). We find no statistically significant differences in survey-reported separations or employer reposting within 30, 60, or 90 days, although hired applicants with immigrant-sounding names are more likely to reapply for another job within 90 days (Appendix Tables F11 and F12). However, these results rely on comparisons between outcomes among hired applicants, not at the hiring margin, and the underlying data could be subject to measurement errors, or the survey sample could suffer selection bias.

[23]Workload is measured using applications received at the reviewer's other postings during the preceding seven days. Appendix F.1 provides the full analysis of vacancy characteristics, missing information, application order, and workload.

Figure 3: Ethnic gaps in interviews and hiring probabilities among all and shortlisted applicants



Notes. This figure reports adjusted interview and hiring probabilities by name group from Equation (1). Panel A uses all applications. Panel B restricts the sample to shortlisted applicants. The two panels use different scales. Vertical lines show 95 percent confidence intervals based on standard errors clustered at the posting level. Appendix Table F1 reports the full estimates.

Age Disparities. Age provides a complementary demographic comparison.[24] Estimating Equation (1) with age in place of G_i, ten additional years are associated with 2.9, 0.36, and 0.21 percentage point lower probabilities of shortlisting, interviewing, and hiring, respectively, all statistically significant (Appendix Table F2). Applicants above the median age are 34 percent less likely to be shortlisted and 53 percent less likely to be hired than younger applicants, gaps comparable to those associated with immigrant-sounding names. The patterns differ among shortlisted applicants: while the name gap in hiring is small and statistically insignificant, the age gap remains negative and significant even if much smaller (Appendix Figure F2). Although conditioning on shortlisting does not isolate employer preferences, this contrast is consistent with age-related disparities persisting in downstream employer decisions.

4.2 Heterogeneity Across Human Reviewers

The average screening gap masks potential differences in how reviewers evaluate applicants. Although instructed to apply the same criteria, reviewers may differ in how they interpret

[24] We find no evidence of a female disadvantage. Female applicants have a small conditional shortlisting premium, while interview and hiring differences among shortlisted applicants are imprecisely estimated (Appendix Figure F2).

applicant information, anticipate employers' preferences, or respond to time and workload constraints. Such heterogeneity matters for AI screening: replacing reviewer discretion with a common algorithmic rule may reduce variation in screening disparities arising from differences in reviewers' judgments. To examine whether shortlisting gaps vary across reviewers, we estimate the following specification, controlling for applicant characteristics and differences in the vacancies they handle:

$$Y_i = \alpha_j + \gamma_{r(i),t} + X_i'\lambda + M_i'\theta + G_i[\beta_{r(i)} + Z_j'\pi] + \varepsilon_i. \tag{2}$$

where $r(i)$ identifies the reviewer and $\kappa_{r(i),t}$ are reviewer-by-month fixed effects that absorb differences in overall selectivity, common to both name groups. The coefficients $\beta_{r(i)}$ allow the conditional name gap to vary across reviewers, while interactions between name classification and observed vacancy characteristics Z_j account for differences in gaps across the types of vacancies reviewers handle. We evaluate each reviewer's gap at a common distribution of Z_j, so differences do not reflect reviewers' job mix. Standard errors are clustered by posting.

Figure 4 shows substantial heterogeneity across eleven reviewers, and we reject equality of reviewer-specific gaps ($p < 0.001$). The figure reports empirical-Bayes estimates, which shrink less precise gaps toward the mean, and posterior 95 percent confidence intervals.[25] The equally weighted mean gap is -3.4 percentage points, with a sampling-error-adjusted standard deviation of 1.8 points. This dispersion is economically large: the implied penalty is 5.2 percentage points for a reviewer with a gap one standard deviation larger than the mean, compared with 1.6 points for one with a gap one standard deviation smaller. Reviewer heterogeneity also persists when measuring relative rather than percentage-point gaps: a Poisson pseudo-maximum likelihood (PPML) specification rejects equality of reviewer-specific name gaps ($p < 0.001$; Appendix Table G1).

We examine the robustness of reviewer heterogeneity in Appendix G. To assess whether reviewer differences persist across samples, we randomly split each reviewer's postings into two halves and reestimate the model. The resulting reviewer-specific gaps have a correlation of 0.82 across the two samples (Appendix Figure G3).

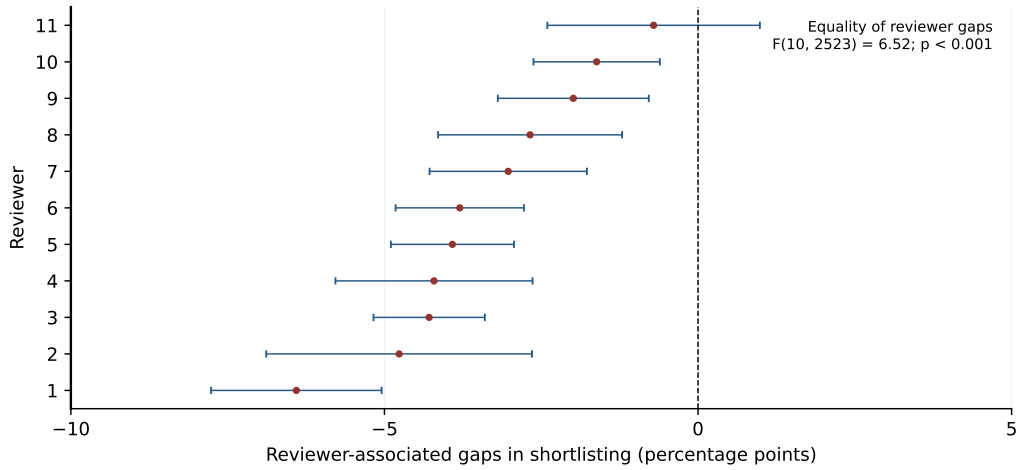
We also examine whether reviewer heterogeneity is associated with workload and prior screening experience, measured by the volume of applications previously handled. Accounting for these factors reduces the sampling-error-adjusted standard deviation of reviewer-specific name gaps only from 1.81 to 1.75 percentage points, and equality of reviewer gaps

[25] Figure G1 reports the corresponding unshrunk estimates with 95 percent confidence intervals based on posting-clustered standard errors.

remains rejected ($p < 0.001$; Appendix Figures G5, G6, and G7).

Together, these exercises show that reviewer-associated differences are stable across sets of postings, and that difference is workload or experience explain little of the dispersion among reviewers.

Figure 4: Reviewer-associated gaps in shortlisting



Notes. This figure reports reviewer-specific shortlisting gaps between applicants with immigrant-sounding and Italian-sounding names from Equation (2). The vertical axis lists anonymized reviewers. The horizontal axis reports gaps in percentage points, evaluated at a common distribution of vacancy characteristics. Points show empirical-Bayes estimates, and horizontal lines show approximate 95 percent posterior intervals. Appendix G describes the estimation procedure.

Reviewer Differences in Applicant Evaluation. We next examine whether reviewer heterogeneity extends beyond perceived origin to other applicant characteristics. We estimate separate shortlisting regressions for each reviewer, allowing the coefficients on applicant and match characteristics to vary (Appendix Table G2). We reject equality of coefficients across reviewers for perceived origin, age, gender, work experience, and language compatibility after adjusting for multiple testing, with particularly large differences in distance compatibility. These patterns indicate that reviewers differ in how they evaluate applicant characteristics, despite being instructed to follow the same criteria.

4.3 Reviewer and employer heterogeneity

Ethnic shortlisting gaps also vary substantially across the employers the reviewers screens applicants for. After accounting for observed applicant and vacancy characteristics, posting fixed effects, and reviewer-by-month differences in overall selection rates, we reject equality

of employer-specific gaps ($p < 0.001$; Appendix Figure G12). Because this specification does not include reviewer-specific name slopes, these differences may partly reflect which reviewers screen each employer's applications. We therefore ask whether reviewer-associated differences persist after accounting for employer-specific gaps, and whether employer-associated differences remain after accounting for reviewers.

we estimate reviewer and employer group slopes jointly:

$$Y_i = \alpha_j + \gamma_{r(i),t} + X_i'\lambda + M_i'\theta + G_i\left[\delta_{r(i)} + \eta_{e(j)} + Z_j'\pi\right] + \varepsilon_i. \tag{3}$$

$r(i)$ identifies the reviewer, and $e(j)$ the employer. δ_r captures the reviewer's conditional name gap, and η_e captures employer conditional name gap . We define all applicant, CV, match, and WhatsApp controls, posting and reviewer-by-month fixed effects, and interactions with vacancy characteristics Z_j as in Equation (2). The only change is that the name gap now includes both reviewer- and employer-specific components. Separating the two sets of slopes requires sufficient within-posting group variation and a connected reviewer–employer network. Employers served by multiple reviewers link reviewers. Reviewers serving multiple employers link firms. Our connected sample includes 36,710 applications to 756 postings, eleven reviewers, and 145 employers.[26] Appendix G.3 reports the corresponding estimates. Reviewer-associated gaps remain heterogeneous after accounting for employer-specific gaps, and employer-associated differences persist after accounting for reviewers (Appendix Figures G9 and G13). Thus, the observed heterogeneity is not captured by either organizational dimension alone.[27] We next study whether reviewer and employer variation differs between ethnic and age screening gaps. To do so, we use a smaller sample in which we can estimate both gaps for the same reviewer–employer pairs.[28] We re-estimate Equation (3), replacing the separate reviewer and employer group slopes with a slope specific to each reviewer–employer pair. The unit of observation remains the application. The ethnic specification controls for age, while the age specification controls for perceived origin. In a second step, we model the reviewer–employer pair estimated gaps as the sum of a reviewer component, an employer component, and a component specific to each reviewer–employer

[26] We keep job postings with both name groups. We also require that reviewers have at least ten postings, 100 applications, and 30 applications from each name group. Employers must have at least two postings and two reviewers, 30 applications, and ten applications from each name group.

[27] A leave-out correction following Kline, Saggio and Sølvsten (2020) applied to the joint model, yields greater employer-associated than reviewer-associated dispersion in the more restricted sample. These estimates are sensitive to sample support (Appendix G.5)

[28] This sample is composed of 16,812 applications, 361 postings, ten reviewers, 42 employers, and 87 reviewer–employer pairs. We keep any pair with at least two postings. Age is measured in ten-year units. Appendix G.4 reports the construction and sensitivity analyses.

pair. Finally, using a crossed random-effects model that accounts for the sampling covariance of the first-stage estimates, we estimate each component's variance. (See Appendix G.4). For ethnic gaps, the estimated standard deviation is 1.69 percentage points across reviewers and 1.21 across employers. For age, the estimated standard deviations are 0.44 and 2.17 percentage points (Appendix Table G5). The point estimates therefore suggest that reviewer-associated variation is relatively more important for ethnic gaps, while employer-associated variation is more important for age. These results motivate examining whether employers respond differently to changes in ethnic and age screening disparities when screening authority shifts to AI. If age disparities in screening reflect employers' preferences, they could reappear downstream even if the AI screener reduces them. If instead ethnic penalties reflect, at least in part, reviewers' own penalties or mistaken expectations about employers, then if the AI screener reduces them we should expect a narrowing in the hiring ethnic gap.

5 How AI Changes the Screening Rule

The preceding section shows that human screening produces penalties that differ across reviewers and employers. In this section, we ask whether a common AI rule changes the shortlisted sample, selection-associated characteristics, and the distribution of screening gaps. We run the AI screener ex post on a subset of applications previously evaluated by human reviewers. Holding the applicant pool and shortlist size fixed within each vacancy isolates differences in selection from changes in screening capacity. This exercise compares screening rules: employers' responses to the counterfactual AI shortlist are not observed.

5.1 Human versus AI selection at equal capacity

For each vacancy, we construct an AI shortlist containing the same number of applications as the observed human shortlist. Let i index applications and $j = j(i)$ the vacancy to which application i was submitted. Y_i^H denotes observed human shortlisting, and K_j is the number of applications shortlisted by humans for vacancy j.

We construct Y_i^A by ranking the same applications by AI score and selecting the top K_j.[29] The comparison covers 19,044 applications to 1,063 vacancies.[30]

[29]Meaning that within each vacancy, applications above the cutoff receive selection probability equal to one and those below the cutoff to zero. When applicants tie at the cutoff, we divide the remaining places equally among them.

[30]The initial pool contains 27,230 applications with observed human decisions, AI scores, and a NamSor name classification. We retain vacancies where humans shortlisted at least one but not all applicants, so selection can differ between the two rules. Each rule selects 2,918 applications in the retained pool.

The two rules select substantially different candidates. Among applicants selected by at least one rule, only 33.1 percent are selected by both; equivalently, AI retains only 49.8 percent of the human shortlist (Appendix Table H1). Selection overlap is lower among applicants with immigrant-sounding names: 28.8 percent of those selected by either rule are selected by both, compared with 35.4 percent among applicants with Italian-sounding names.

5.2 How do human and AI screening criteria differ?

To examine how selection relates to applicant characteristics, we estimate parallel regressions on the same applications:

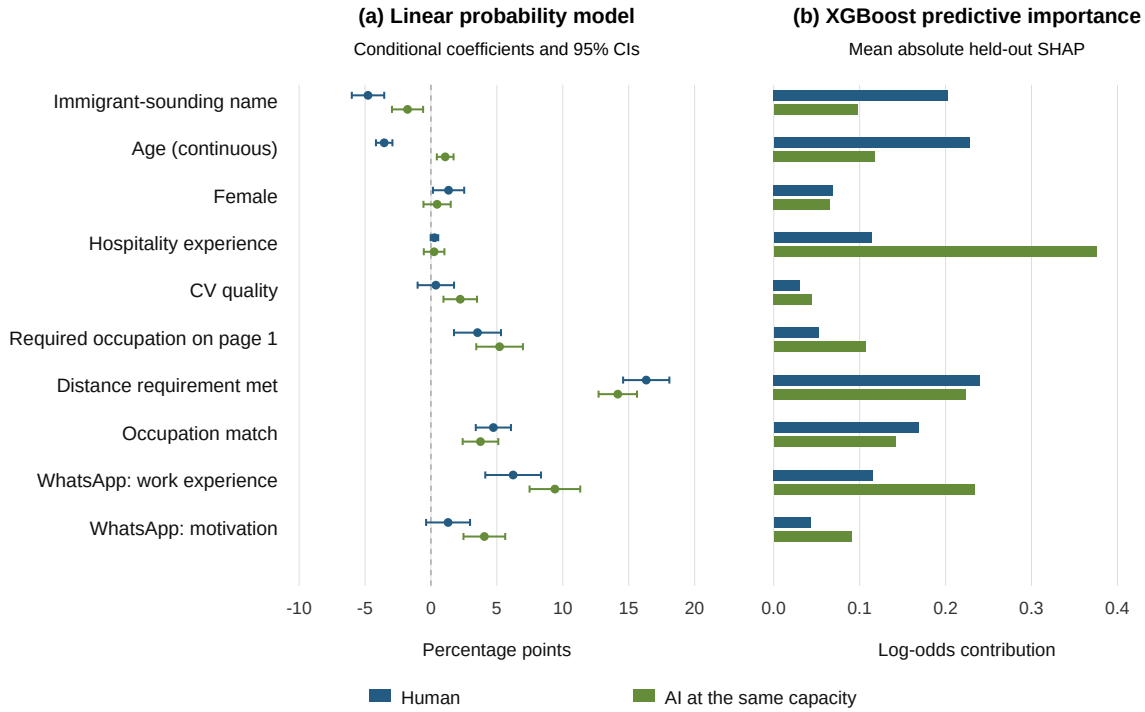
$$Y_i^s = \alpha_j^s + \lambda_t^s + \beta_s G_i + X_i'\lambda_s + M_i'\theta_s + \varepsilon_i^s, \qquad s \in \{H, A\}. \tag{4}$$

G_i is the indicator for immigrant-sounding name, X_i contains applicant and CV characteristics, and M_i contains candidate–vacancy match and WhatsApp measures. Posting fixed effects α_j^s absorb vacancy-specific selection levels, and λ_t^s denotes month–year fixed effects. We test differences directly using the within-application outcome difference $D_i = Y_i^A - Y_i^H$, with the same regressors and posting-clustered standard errors (Appendix H). These comparisons describe conditional selection gaps.

Figure 5, panel (a), shows that the conditional ethnic shortlisting penalty falls to -1.78 percentage points under AI, a reduction of 63 percent. The difference between the two rules is statistically significant. In addition, in this linear specification, the significant penalty associated with older age under human screening is also absent (reversed) under AI. The smaller conditional ethnic gap under AI also remains under flexible adjustment using XGBoost, both with pooled residualization and with a within-posting paired comparison (Appendix Table H2). If we allow for non linearities, still the age-related disadvantage remains significantly smaller under AI with respect to humans, even if they both follow a reversed U-shape (Appendix Figures H1 and H3).

Panel (b) complements these linear estimates by examining which applicant characteristics are most informative in predicting selection under each rule. We fit separate XGBoost models that allow for nonlinearities and interactions and summarize predictive importance using mean absolute SHAP values. Hospitality experience has greater predictive importance under AI, whereas age and immigrant-sounding names have greater predictive importance under human screening. These measures capture how strongly each characteristic contributes to

Figure 5: Human and AI screening: conditional associations and predictive importance



Notes. Both panels use the paired pool of 19,044 applications. Panel (a): separate linear regressions with posting and month–year fixed effects and the full preferred controls. Binary coefficients are $100\widehat{\beta}$ and continuous coefficients are $100\widehat{\beta}SD(X)$. 95 percent confidence intervals clustered by posting. Age is continuous. Panel (b): it shows mean absolute held-out SHAP contributions in log odds from separate XGBoost models using the same applicant and match controls. Five-fold cross-fitting holds out entire postings. Contributions of each variable and its missing indicator are summed before taking absolute values. Additional importance rankings and nonlinear profiles under the earlier control specification appear in Appendix Figures H2 and H3.

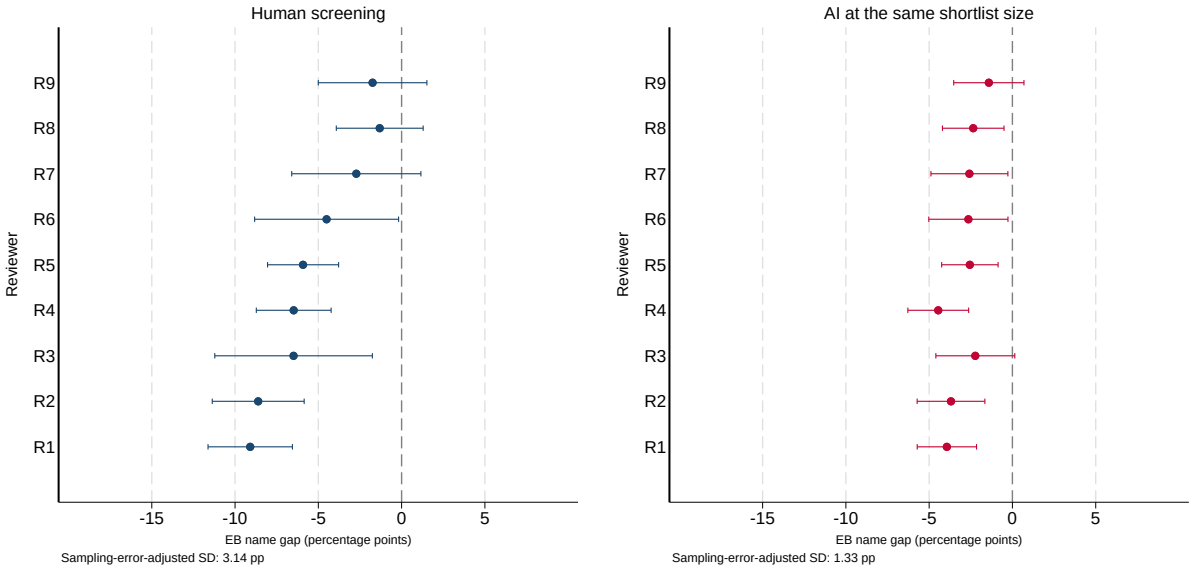
the fitted models' predictions, not whether it increases or decreases selection. They describe predictive associations rather than the production algorithm's internal evaluation criteria.

Together, these results show that, even when selecting the same number of applicants from the same pool, humans and AI differ in conditional ethnic and age disparities, as well as in the predictive importance of applicant characteristics.

5.3 Does a common AI rule reduce variation across reviewer applicant pools?

We next ask whether the change is limited to the average ethnic gap (*the level*) or also affects its *dispersion* across the applicant pools previously screened by different reviewers. We estimate reviewer-associated name gaps for each rule on the same applications, allowing gaps to vary with observed vacancy characteristics and evaluating them at a common vacancy

Figure 6: Ethnic gaps across reviewer applicant pools under Human and AI screening



Notes. This figure reports shortlisting gaps between applicants with immigrant-sounding and Italian-sounding names across reviewer applicant pools. Both panels use the same 18,685 applications to 1,001 postings assigned to nine reviewers, with the same controls. The vertical axis lists anonymized reviewers, with the same labels and ordering in both panels. The horizontal axis reports gaps in percentage points. Points show empirical-Bayes estimates, and horizontal lines show approximate 95 percent posterior intervals. The mean gap changes from −5.15 under Human screening to −2.70 under AI screening. The standard deviation adjusted for sampling noise falls from 3.14 to 1.33 percentage points.

distribution. Posting and reviewer-by-month fixed effects account for differences in overall selection. Appendix H.4 gives the specification, controls, support restrictions, and correction for estimation noise. Under AI, reviewer identity labels the original applicant pool, not a different AI decision-maker.

Figure 6 shows that AI reduces not only the average conditional ethnic shortlisting gap but also its estimated dispersion across reviewer applicant pools. Thus, the change in screening technology affects both the level of disparities and how unevenly they are distributed across applicant pools.

A common rule does not need to remove all the variation across applicant pools. Still there could be different group gaps depending on who applies and on the vacancy characteristics. Also, the common rule could still be biased.

The estimated association between workload and the ethnic screening penalty is about 65 percent smaller under AI than under human screening, though the difference is estimated imprecisely ($p = 0.676$) (Appendix Table H4).

At fixed capacity, the AI screener changes the selected candidates, the characteristics associated with selection, and the estimated distribution of ethnic gaps. Whether these

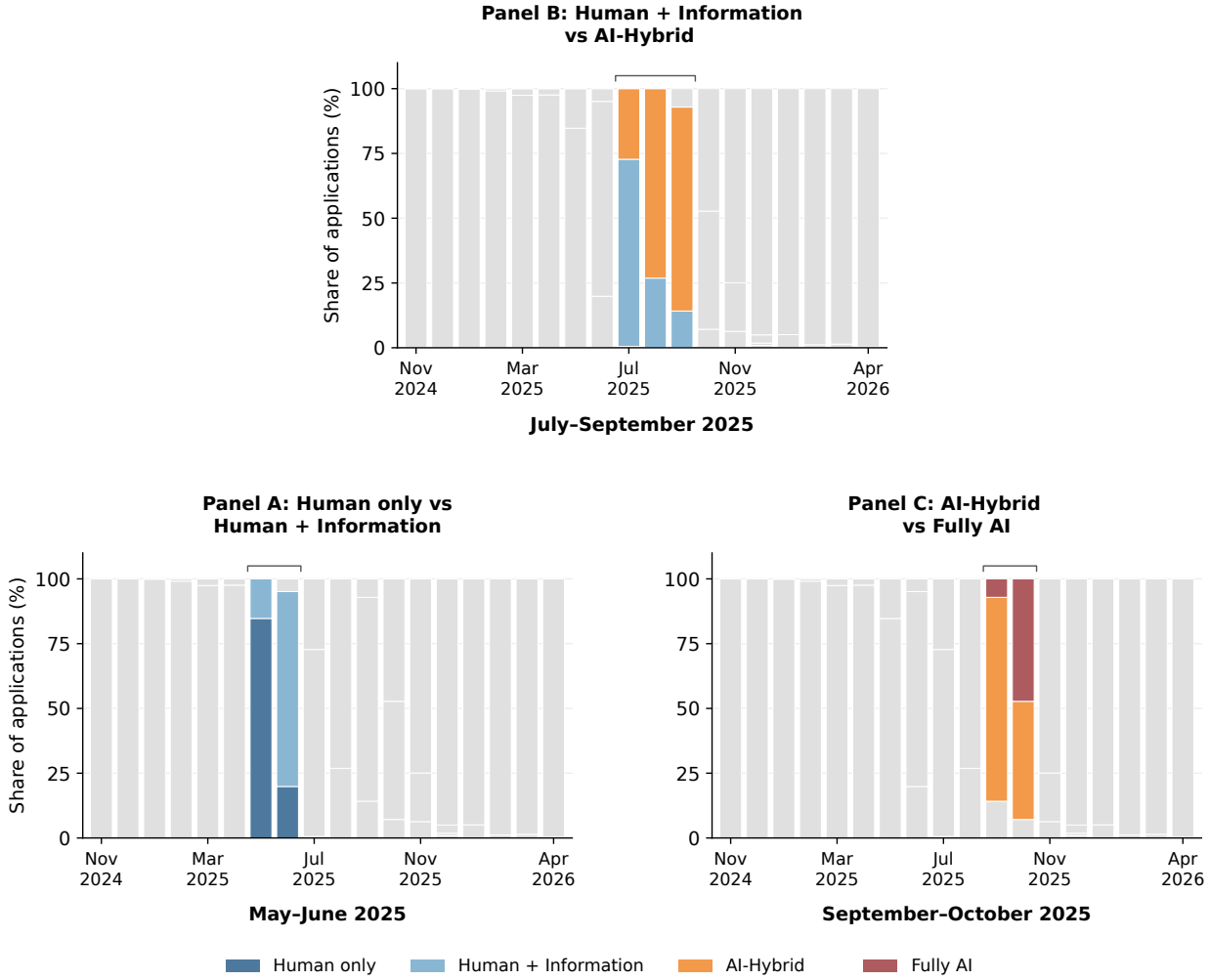
changes improve access to employment or matching depends on how employers respond to the resulting choice sets, which we examine using the randomized deployment.

6 The effects of AI screening on employer decisions

So far, we have shown that when human reviewers held authority over screening, there are economically meaningful ethnic and age penalties, that these disparities vary across reviewers and employers, and that they do not disappear even when we take into account vacancy characteristics, workload, and attention effects. We have also shown that, for the same vacancies and applications, AI screener imposes a significantly smaller ethnic penalty, eliminates or reverses the age penalty, and reduces dispersion across reviewers. In this section, we study what happens when employers receive different shortlisted sets. In particular, we examine whether screening changes translate into changes in interviews and hiring, and whether employers reinforce or offset the reduction in disparities at the screening stage.

As illustrated in Figure 2 and Figure 7, starting in May 2025, the platform gradually transitioned to AI screening, with different regimes operating simultaneously during the transition. As Panel A of Figure 7 shows, the platform first moved from the *Human-only regime* to a regime in which human reviewers still had full decision-making authority, but they received additional information: the scores produced by the AI screener (*Human + Info regime*). Then, as shown in Panel B, the platform introduced the *AI-Hybrid regime*, in which all applications scored above 8 by the AI were automatically sent, and all below 5 were automatically rejected. Those between 5 and 8 could be reviewed by humans. Finally, Panel C shows the introduction of the *Full AI regime*, in which all applications scoring above 5 were automatically sent to employers.

Figure 7: Screening regimes and deployment comparisons



Notes. The horizontal axis shows application month; the vertical axis shows the percentage of applications under each screening regime. The three panels show the same monthly distribution. We highlight the two regimes and months used in each comparison: Human only versus Human + Information in May–June 2025, Human + Information versus AI-Hybrid in July–September 2025, and AI-Hybrid versus Fully AI in September–October 2025. Other regimes and months are shown in gray.

6.1 Empirical strategy

For all overlapping comparisons shown in Figure 7, we run the same main analysis described below. These comparisons cover different periods and populations, so we interpret each separately. During May-June 2025 (Panel A), the platform started testing AI and the AI scores were visible to human reviewers who retained full decision-making authority. From July 2025, the platform started testing the delegation of screening to AI. According to the platform, it assigned screening configurations at the job-posting level without any specific

criteria, generating a good as random variation in screening regimes.[31]

For each comparison, our main estimating equation is the following:

$$Y_i = \alpha + \tau T_j + \beta G_i + \delta(T_j \times G_i) + X_i'\gamma + M_i'\theta + \lambda_t + \varepsilon_i, \tag{5}$$

where i indexes applications, j the job posting, and t the month of application. Y_i is an indicator for the outcome of interest: whether the applicant's profile is opened, the applicant is interviewed, or the applicant is hired. G_i is an indicator for an immigrant-sounding name. $T_j = 1$ denotes assignment to the regime with greater use of AI in the relevant comparison. X_i contains applicant and CV characteristics, M_i contains candidate–job match and WhatsApp controls, and λ_t denotes month fixed effects. The coefficient τ is the treatment difference for applicants with Italian-sounding names, $\tau + \delta$ is the corresponding difference for applicants with immigrant-sounding names, and δ measures the change in the name gap. We cluster standard errors at the employer level. A causal interpretation requires regime assignment to be independent of potential outcomes, conditional on the included controls and month fixed effects. We interpret the estimates as Intention-to-Treat effects.

For each comparison, we report a balance table describing the job postings included in the pooled regressions estimated using Equation (5). As an additional robustness check, we use coarsened exact matching (CEM) to compare postings with similar observed characteristics across regimes.

We distinguish changes in disparities from changes in total hiring. At the posting level, we examine whether employers interview or hire anyone, along with screening time. We then examine employer satisfaction and measures of match stability, including reported separations.

We start by providing the main results for the comparison between the **Human regime**, in which reviewers still had full authority over shortlisting, and the **AI-Hybrid regime** as described above, in which the authority for the majority of applications shifted to the AI screener.[32] We focus on this comparison because this transition captures the main technological change or shift in authority. After describing the main results in this section, we examine the mechanisms in Section 7, while in Section 8 we provide results for the other comparisons between screening regimes.

[31] This description of assignment is based on information provided by the platform. We assess balance in observed characteristics and run the robustness checks described below to test how sensitive our estimates are to differences across regimes.

[32] It is actually Human + Information, but for the sake of exposition we use them interchangeably here

6.2 Main results: *Human* versus *AI-Hybrid*

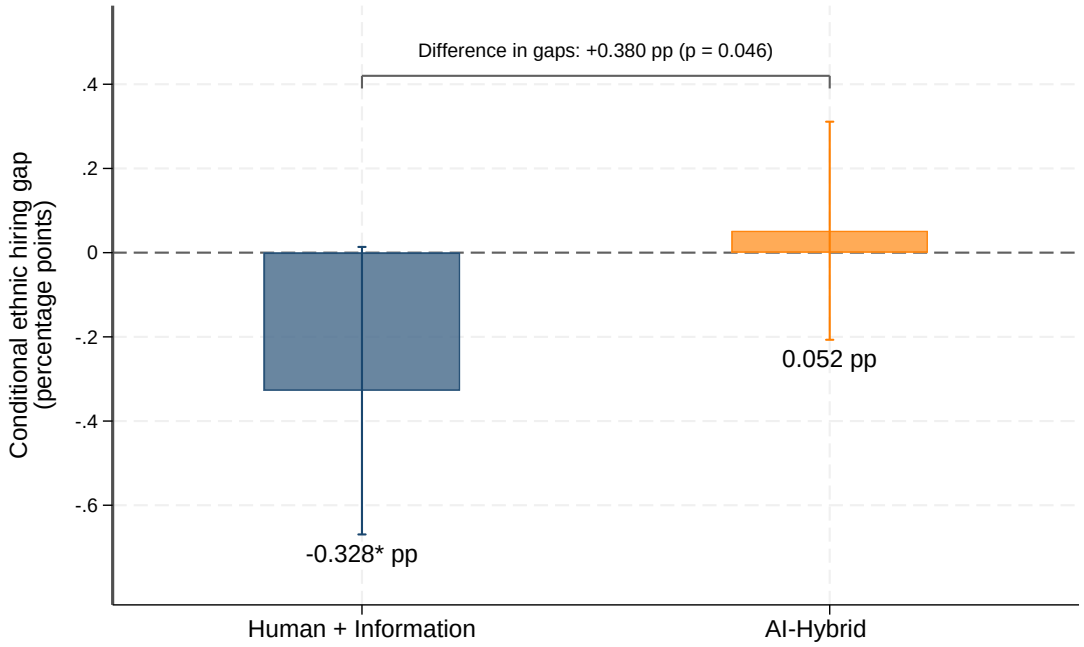
We study these regimes during their overlap period from July through September 2025; our main sample covers 210 job postings assigned to Human and 393 assigned to AI-Hybrid.

In Appendix Table I1, we compare predetermined posting, employer, and applicant-pool characteristics between postings assigned to the two regimes. Among the twenty-three characteristics examined, only three (seasonality, experience requirements, and days per week) are significantly different between the two regimes, and only one (days per week) has an absolute standardized difference above 0.25. In the CEM analysis, we therefore match job postings on these characteristics and obtain absolute standardized differences below 0.25 for all characteristics. We also conduct additional robustness checks: the estimated changes in outcomes remain in the same direction across specifications. However, their magnitudes differ slightly, partly because the estimation sample can change.

Hiring and ethnic disparities. Figure 8 shows that *AI-Hybrid* closes the conditional hiring gap between applications with immigrant and Italian-sounding names. The figure reports the estimated hiring gap under Human+ Information (β), the gap under AI-Hybrid ($\beta + \delta$), and the change between regimes (δ) from Equation (5). Under Human + Information, applicants with immigrant-sounding names have a 45 percent lower adjusted hiring probability than applicants with Italian-sounding names. Under AI-Hybrid, this disadvantage disappears in the point estimates: the conditional gap changes from -0.328 to $+0.052$ percentage points, an improvement of 0.380 points ($p = 0.046$). The adjusted hiring probability for applicants with immigrant-sounding names rises from 0.401 to 0.615 percent, an increase of approximately 53 percent ($+0.214$ percentage points; $p = 0.074$). Appendix Table I2 reports the regression coefficients.

Appendix Table I3 instead shows results from several *robustness checks* that let us test whether this narrowing in hiring probabilities depends on vacancy characteristics, their relationship with ethnic gaps, and reviewer assignments. First, we allow the ethnic gap to vary with new employer status, seasonality, minimum experience requirements, and days per week (jointly and separately), then we introduce job-posting fixed effects that allow us to compare applicants within the same vacancy and absorb differences in baseline hiring probabilities across postings; finally, we allow for reviewer-by-month fixed effects to account for possible differences in baseline hiring probabilities across reviewer assignments within each month. Across all specifications, δ remains between 0.36 and 0.40 percentage points and is statistically significant.

Figure 8: AI-Hybrid versus Human + Information: the conditional hiring gap



Notes. This figure reports conditional hiring gaps between applicants with immigrant-sounding and Italian-sounding names from Equation (5). The sample contains 31,198 applications to 603 postings in July–September 2025. The horizontal axis shows the screening regime. The vertical axis reports hiring probability gaps in percentage points. Bars report β under Human + Information and $\beta + \delta$ under AI-Hybrid. Vertical lines show 95 percent confidence intervals. Standard errors are clustered at the employer level. Appendix Table I2 reports the regression coefficients. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix Table I2 shows the results also for the matching procedure (CEM), in which on the 353 postings retained after matching each application receive its posting' CEM weights, and we use the same controls as in the main specification. The estimated adjusted hiring gap narrows by 94 percent (0.666 percentage points, $p = 0.043$). Appendix Table I3 provides additional robustness checks for the matching procedure leading to similar results.

This stable reduction in ethnic hiring gaps is consistent with reviewers imposing their own group penalties or incorrectly anticipating employer biased preferences. We examine the mechanisms behind these changes in Section 7.

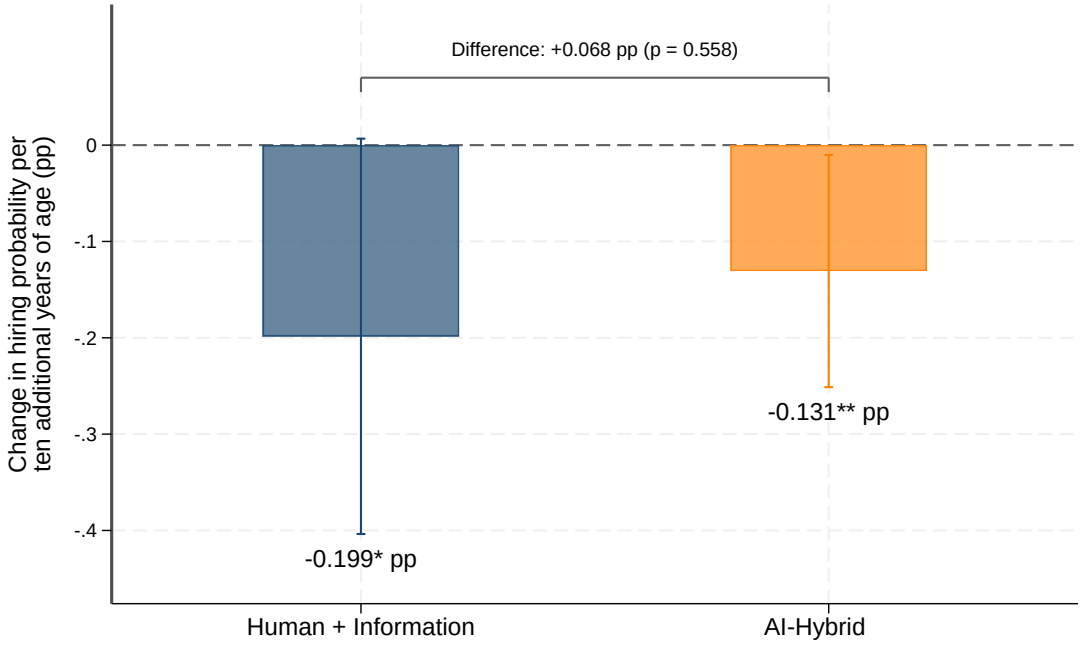
For interviews, we find no detectable change in the ethnic gap across regimes (Appendix Figure I1). The estimated change is +0.054 percentage points in the pooled sample ($p = 0.871$) and +0.325 points in the CEM-weighted matched sample ($p = 0.632$; Appendix Table I2).

Age and the interpretation of convergence. Figure 9 shows that the same results do not hold for age even though at screening older applicants gain access under AI: *AI-*

Hybrid does not eliminate the negative conditional relationship between age and hiring. The figure reports the estimated hiring gap under Human+ Information (β), the gap under AI-Hybrid ($\beta + \delta$), and the change between regimes (δ) from Equation (5) using age in ten-year units. Under Human + Information, ten additional years of age are associated with a 0.20 percentage point lower hiring probability. Under AI-Hybrid, the corresponding estimate is 0.13 percentage points. Even if there is a small attenuation, the difference is not statistically significant ($p = 0.558$). Appendix Table I4 reports the regression coefficients. These results are robust when we use the matching sample (Appendix Table I4). For interviews, older applicants also remain disadvantaged under both regimes.

This pattern is consistent with the earlier analysis of reviewer and employer heterogeneity (Section 4.3) and the simple framework Section K i.e. with reviewers correctly anticipating employer preferences and with employers reintroducing the age penalty.

Figure 9: AI-Hybrid versus Human + Information: age and hiring probabilities



Notes. This figure reports coefficients from the age specification of Equation (5). The sample contains 26,047 applications to 602 postings in July–September 2025. The horizontal axis shows the screening regime. The vertical axis reports the change in hiring probability associated with ten additional years of age, in percentage points. Vertical lines show 95 percent confidence. Standard errors are clustered at the employer level. Appendix Table I4 reports the regression coefficients. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Posting outcomes and screening time. We now analyze posting-level outcomes in Table 2 for the same pooled sample of 603 postings. AI-Hybrid increases the probability

of at least one interview by 38 percent (11.5 percentage points) relative to the Human + Information mean of 30 percent. Instead, we do not find any significant change in the probability of at least one hire: it increases by 15 percent relative to the control mean of 21 percent.[33] At the same time, average shortlisting time falls by 1.33 days, not surprisingly, since AI decisions are instantaneous, and we find no statistical difference in how long a job posting stays online. Taken together, these additional results show that the narrower ethnic hiring gap comes with a slight improvement in efficiency measures (days to shortlisting), without a detectable change in overall hiring.

Table 2: Posting outcomes and employer satisfaction under AI-Hybrid

Outcome	Any interview	Any immigrant-sounding name interview	Any hire	Days to shortlist	Days online	Employer satisfaction index (0–10)
AI-Hybrid	0.115**	0.090**	0.031	−1.325**	−0.472	1.547***
	(0.051)	(0.036)	(0.039)	(0.527)	(0.881)	(0.481)
Control mean	0.300	0.118	0.210	4.322	2.129	5.948
Observations	603	577	603	582	603	212
R^2	0.169	0.188	0.120	0.120	0.059	0.311

Notes. This table reports posting-level regression estimates for July–September 2025. In Column (1), the dependent variable is an indicator for whether the employer interviewed at least one applicant. In Column (2), the dependent variable is an indicator for whether the employer interviewed at least one applicant with an immigrant-sounding name. In Column (3), the dependent variable is an indicator for whether the employer hired at least one applicant. In Column (4), the dependent variable is days to shortlisting. In Column (5), the dependent variable is days online. In Column (6), the dependent variable is the employer satisfaction index. The employer satisfaction index is calculated as the equally weighted mean of satisfaction with the service, satisfaction with the assistance received, and willingness to recommend the platform, each measured from 0 to 10. Standard errors are clustered at the employer level. Appendix Table I5 reports the satisfaction components and willingness to reactivate the service. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Employer satisfaction and retention. *Employer satisfaction.* Importantly, Table 2 also shows that employers under the AI-Hybrid regime report higher satisfaction with the job-matching platform. In particular, the satisfaction index, which combines satisfaction with the service, satisfaction with the assistance received, willingness to recommend the platform, and willingness to reactivate the service, increases by 26 percent relative to the Human+ info mean. Appendix Table I5 shows the index components separately. Although there could be selection bias in responses, we think it is unlikely to differ between regimes,

[33]The dependent variable is equal to 1 if hiring happened through the platform, 0 otherwise. The baseline percentage is normal in Italy, where informal recruitment channels remain the most common source of employment: 41.2 percent of firms use personally known candidates and 34.2 percent use friends or relatives.(Unioncamere and Ministero del Lavoro e delle Politiche Sociali, 2025)

given that employers did not know about the AI screener. These results are important, showing that the narrowing gap does not come at the cost of lower satisfaction, even though higher satisfaction could reflect faster screening, preferences for longer lists, differences in the candidates received, or all of them.[34]

Retention. We also do not find meaningful differences in separation rates across three different measures: employment separations reported by employers in the survey, subsequent applications by hired workers, and employer repostings in the same occupation. Appendix Table I6 shows these results. Each measure has its own limitation: (i) the sample of the survey outcomes is small and could be selected, even though we think it is unlikely to differ between regimes and the sounding name of the hired person; (ii) subsequent applications by hired workers or employers reposting could be subject to measurement errors.[35] However, the evidence is quite consistent with the idea of not worse outcomes.

Taken together, we have shown that AI-Hybrid narrows the ethnic hiring gap, speeds up screening, and increases satisfaction among responding employers, without a detectable reduction in overall hiring or matching stability. The pattern differs for age: older applicants remain less likely to be hired, despite the sizeable disparity reduction at screening. We now investigate how AI screening changes shortlist size, the characteristics of shortlisted applicants, and ordering, and how employers respond to the candidates they receive.

7 Mechanisms

Why does AI screening narrow the ethnic hiring gap? In this section, we investigate whether hiring changes come from better ranking of shortlisted applicants (ranking), greater access to employers (size), or changes in the shortlisted set's characteristics (applicant characteristics and job fit). These mechanisms also help explain how a common screening rule more generally shapes employer decisions.

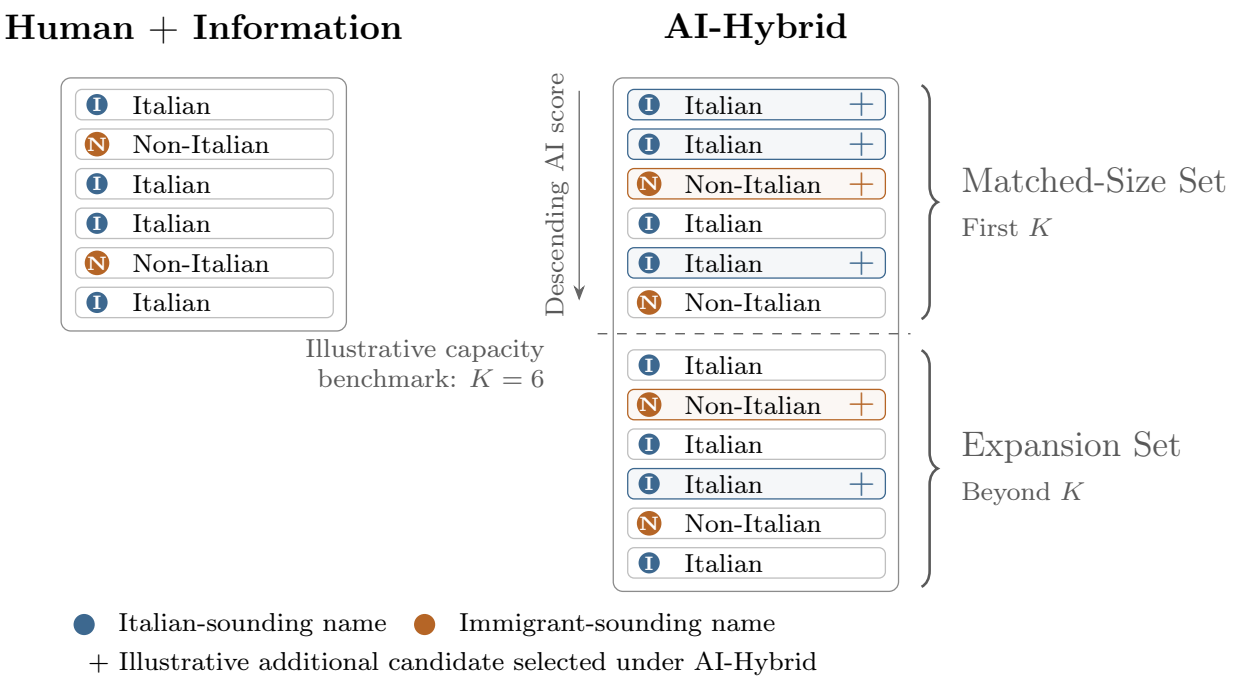
Figure 10 illustrates how the shortlisted set actually changed on average between the two regimes. As explained below, the size of the shortlisted pool increased, and the ordering also changed, with the AI shortlisted set ordered by descending AI score (highest to lowest).[36] In the following analysis, we refer to the matched-size set when we keep the set size constant, and to the expansion set when we refer to applications beyond the matched-size set.

[34] We explain this in Section 7

[35] We are working with the Job Matching Platform to increase this.

[36] In the human selected shortlisted set the ordering depended only on the time of arrival and evaluation.

Figure 10: How screening changes employers' choice sets



Notes. This figure illustrates a schematic example of how the choice sets of employers would appear under the two regimes. The choice set under the AI-Hybrid regime (right panel) is different in the number, characteristics, and ranking of applicants. Blue cards represent applicants with Italian-sounding names, and orange cards represent applicants with immigrant-sounding names. The + marks additional illustrative candidates under AI-Hybrid. The matched-size set contains the six highest-scoring AI-shortlisted candidates, and the expansion set contains the remaining candidates. In the empirical analysis, the capacity benchmark varies across CEM strata and is based on the mean shortlist size under Human + Information within the same stratum.

7.1 Ranking and employer attention

Do we see a reduction in ethnic hiring gaps because immigrants appear closer to the top of the list? Employers may have limited attention, so narrowing ethnic hiring gaps could simply reflect better placement. We find no statistically detectable improvement in the relative ranking of applicants with immigrant-sounding names, with consistent results across the pooled and CEM comparisons (Appendix Table I8).

Moreover, when we control for rank and AI scores among shortlisted applications, the estimated reduction in the ethnic hiring gap is largely unchanged. Results are consistent across the pooled and CEM comparisons (Appendix Tables I9 and I10).

Finally, we do not find an overall decrease in attention: employers actually open more profiles under AI-Hybrid screening (Appendix Table I11), and this does not change by immigrant-sounding name (Appendix Tables I9 and I10). Together, these findings do not find sufficient evidence that improved relative ranking or a reallocation of attention toward applicants with immigrant-sounding names is the main explanation for the narrowing of the ethnic hiring gap.

7.2 Access: shortlist size and use of AI scores

Does AI give immigrant applicants more access to employers by expanding the shortlist? Larger shortlists could allow more applicants with immigrant-sounding names to reach employers, even without improving their relative ranking. We find that AI-Hybrid increases shortlist size by 5.87 candidates per posting, 70 percent of whom have Italian-sounding names (Appendix Table I12). Therefore, immigrants gain in absolute terms but not relatively.

Importantly, in both name groups, two-thirds of the added applications come from applicants scoring at least 8 under AI, and therefore positioned at the top of the list (Table I12 and Figure I3). This means employers now receive more candidates, and those favored by AI but not included by humans are placed at the top of the list; i.e., in Figure 10, they enter the matched-size set.

7.3 Applicant characteristics and job fit at comparable screening capacity

Does AI change the characteristics and job fit of applicants who reach employers, even at comparable shortlist size? Larger shortlists may include different types of candidates simply

because they include more applicants. Under AI-Hybrid, in the complete shortlist, consistent with what we have found in Section 5, applicants are older, and applicants with immigrant-sounding names have a higher composite match index (Appendix Table I13). This already points to different applicant characteristics and job fit, which allows "better" immigrants to reach the final employers.

We therefore ask whether differences in applicant characteristics and job fit remain when we compare shortlists with similar screening capacity. To do this, we use the CEM sample, which lets us use the average shortlist size among human-matched vacancies as a benchmark. Therefore, for each AI-Hybrid job posting, we create a matched-size set that matches the benchmark size using the highest-scoring AI-shortlisted candidates, and an expansion set that includes shortlisted applications beyond the matched-size set (Appendix I.1).

Table 3 shows that, even at comparable shortlisted capacity, under AI-Hybrid both name groups are older and have more hospitality experience. The composite match index increases overall, indicating possibly no deterioration in quality and even an increase, although the positive differences within each name group are imprecisely estimated.[37] In particular, applicants with immigrant-sounding names are also less likely to have employment gaps, a possible proxy for turnover that employers care about, and Italians have lower distance match.

Together, these findings suggest that changes in applicant characteristics and job fit go beyond the increase in shortlist size. In particular, applicants with immigrant-sounding names reach employers with more hospitality experience and fewer employment gaps, without a detectable reduction in composite applicant–job fit. These difference observed characteristics could explain the narrowing in ethnic gaps in hiring probabilities. In order to separate out size from applicant characteristics and job fit, we next study where hires come from within the shortlist and whether we observe a narrowing in the ethnic hiring gap keeping the shortlisted size fixed.

7.4 What accounts for the narrowing of the ethnic hiring gap?

To distinguish the effect of size from differences in the applicant pool, we run two exercises. First, using the CEM job postings sample, we check whether the change in the application-level hiring gap comes from the matched-size set or the expansion set (Appendix Table I19). We find that the reduction in the ethnic gap is concentrated in hiring from the matched-size set: it contributes 0.79 percentage points, while the expansion set contributes -0.07 points.

[37] Appendix Table I15 reports the complete set of applicant characteristics and job-fit outcomes. Appendix Table I14 reports applicant characteristics and job fit for the complete shortlist on the same matched support.

Table 3: Applicant characteristics and job fit at comparable screening capacity

	Age	Female	Match composite	Occupation match	Distance match	Hours match	Hospitality exp. (years)	Minimum exp. met	Employment gaps
Italian sounding names									
AI-Hybrid	3.849***	−0.021	0.025	0.023	−0.050*	−0.102	1.550**	0.141***	−0.006
	(1.407)	(0.048)	(0.021)	(0.058)	(0.028)	(0.098)	(0.738)	(0.053)	(0.054)
Control mean	34.265	0.329	0.583	0.536	0.948	0.926	6.000	0.668	0.585
Immigrant sounding names									
AI-Hybrid	3.033**	−0.056	0.035	0.111	−0.014	−0.027	1.593**	0.045	−0.155**
	(1.442)	(0.076)	(0.023)	(0.078)	(0.054)	(0.242)	(0.767)	(0.071)	(0.078)
Control mean	32.364	0.359	0.579	0.543	0.860	0.960	4.475	0.690	0.635
Total matched postings: 347									

Notes. Coefficients compare AI-Hybrid with Human + Information using CEM weights, month fixed effects, and predetermined posting controls. Outcomes are posting-level means within each name group. The AI matched-size set contains the highest-scoring shortlisted applicants up to the rounded mean human shortlist size in the same CEM stratum; shorter lists enter in full and cutoff ties receive fractional membership. Human postings contribute their complete shortlist. The matched sample contains 347 postings; outcome-specific samples depend on the presence of shortlisted applicants in each name group and on available characteristics. Age and hospitality experience are in years; other outcomes are shares or match indices. Standard errors clustered by employer are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Employers therefore hire from the expansion set, but these hires do not account for the estimated narrowing of the ethnic gap.

Second, we study whether the narrowing of the ethnic hiring gap is explained by differences in shortlist size . We reweight AI-Hybrid postings toward the shortlist-size distribution observed under Human + Information.[38] We re-estimate Equation (5) on the same 12,724 applications to 346 postings. The result in hiring ethnic gap does not change (Appendix Table I24), suggesting that differences in shortlist size alone do not appear to drive the ethnic hiring result.

7.5 Does the expanded shortlist provide additional matching opportunities?

We now ask whether the expansion set helps sustain overall hiring, even if it does not explain the narrowing of the ethnic gap. Descriptively, at the posting level, we compare the probability of hiring from the matched-size set versus the expansion set, and we find that employers hire from the expansion set in both name groups (Appendix Table I17). Employers would not hire these candidates if the size were held constant. We do not know what would

[38]We estimate a posting-level model for the probability of belonging to AI-Hybrid using log(1 + shortlist size) and its square, with CEM weights. We then multiply each AI-Hybrid posting's CEM weight by $(1-\widehat{p}_j)/\widehat{p}_j$, where $\widehat{p}_j$ is the fitted probability.

have happened otherwise, but the fact that employers hire from the expansion set (Appendix Table I18) already suggests a possible mismatch between some of the AI screener's evaluation criteria and the employers'.

In particular, the expansion set seems to sustain the age gap we have found before. In the CEM and common sample used to compare ethnicity and age, the narrowing of the ethnic hiring gap is concentrated in hiring from the matched-size set, whereas the estimated change toward a more negative age gradient is concentrated in the expansion set (Appendix Table I22).

We finally examine whether the reduction in ethnic hiring gap comes from immigrants being younger. We compare specifications without age controls, with linear and quadratic age controls, and with five-year age groups, also allowing their coefficients to differ across screening regimes (Appendix Table I23). The estimated narrowing of the ethnic hiring gap remains very similar across these specifications.

8 Other comparisons between screening regimes

We now analyze the other comparisons across regimes in the deployment period in the overlapping samples. Each comparison uses a different period and population, so we interpret these separately. We begin by comparing job postings in the time window May-June 2025 assigned to the **Human only regime** with the ones assigned to the **Human + Information** regime. This comparison examines how providing human reviewers with the score generated by the AI screener changes screening and downstream employers' decisions. Then we move to time-window September–October 2025 in which the comparison is between the **AI-Hybrid** versus **Fully AI** regimes, this comparison asks what happens when we eliminate completely humans even from the intermediate cases.

8.1 *Human only* versus *Human + Information*

This comparison asks what happens to ethnic gaps when human reviewers receive as additional information the scores produced by the AI screener, compared with when this information is not available. We repeat the analysis used for the main comparison (Appendix Table J2). The pooled sample includes 699 job postings, while the coarsened exact matching sample includes 484 postings.

We find no statistically significant or sizeable differences in ethnic gaps in interview or hiring probabilities, and the point estimates are small. The results are similar for age gaps.

(Appendix Table J3). We also do not observe any statistical difference in posting outcomes (any interview, or any hire) and employer satisfaction. However, shortlisting seems to take longer (Appendix Table J4), consistent with the fact that reviewers also had to process additional information.

Finally, we find that reviewers' shortlisting decisions do not fully align with the suggestion given by the AI scores: even among applications scored higher than 8 by the AI screener, only less than 50% are forwarded to employers (Appendix Figure J1).[39] This finding is consistent with what previous research (Hoffman et al., 2018; Awuah et al., 2026) has documented. At comparable shortlist sizes, applicants with immigrant-sounding names selected under Human + Information have a higher composite match index and are more likely to meet the minimum experience requirement than those selected under Human only (Appendix Table J8). Providing AI information therefore changes some observed dimensions of applicant–job fit mostly for immigrants consistent with reducing uncertainty. However still, applicants with immigrant-sounding names have lower shortlisting rates in every recorded AI-score band.

Taken together, these findings suggest that making AI information available to human reviewers is not sufficient to change disparities in the hiring pipeline. Although, we cannot distinguish whether the results are led by a common distrust toward AI or limited attention, applicants with immigrant sounding names remains to be less likely to be shortlisted within the same AI-score band but providing AI information seem somewhat favorable for them. This is consistent, once again, with both reviewers imposing their own group penalties or anticipating employers' preferences.

8.2 *AI-Hybrid* versus *Fully AI*

This comparison asks what happens to ethnic gaps when the AI screener has full authority, meaning it decides even for applications between 5 and , and 8. We repeat the analysis used for the main comparison (Appendix Table J10). The pooled sample includes 464 job postings, while the coarsened exact matching sample includes 400 postings.

We find no statistically significant or sizeable differences in ethnic gaps in interview or hiring probabilities, and the point estimates are small. The results are similar for age gaps (Appendix Table J11). We also do not observe any statistical difference in posting outcomes (any interview or any hire) (Appendix Table J12), and employer satisfaction does not change significantly among respondents (Appendix Table J13).

[39]The comparison with scored applications under Human only provides an additional descriptive benchmark. However, this comparison at the moment is still incomplete (Appendix Figure J2).

Fully AI by construction increases the number of shortlisted candidates. On average, the shortlist size increases by 13.00 candidates per posting, doubling the AI-Hybrid regime. We do not see a decrease in attention: employers open almost 6 more profiles, but attention does not increase proportionally with shortlist size (Appendix Table J16).

Importantly, we examine human selection in the intermediate score band. Among candidates with immigrant-sounding names, Humans select more strongly on occupation and experience, while the overall matching index is not statistically different (Appendix Table J17).

9 Discussion

Our results have three fundamental implications. First, hiring inequalities depend both on how screening authority is allocated and how employers respond to their available choice sets. Second, substituting heterogeneous human discretion with a common AI screening rule can change both how we think about the form of disparities and the stages at which they arise or persist; this has implications for where interventions can be effective. Third, providing humans with AI information and delegating decision authority to AI are distinct organizational choices. This speaks directly to the debate on human-in-the-loop regulation.

Through our natural experiment, we show that partial delegation to AI screening narrows ethnic disparities in hiring, while older applicants remain less likely to be hired. These results are consistent with intermediate reviewers imposing their own penalties and, in some cases, incorrectly anticipating employers' preferences, but not in others. Reducing disparities at the screening stage has the potential to change downstream outcomes without deterioration in employer satisfaction or match stability. In our context, employers did not know that AI constructed the choice sets, which allowed us to make a clear causal comparison based only on the different choice sets presented. However, future research could study what happens when employers are informed that AI constructed the shortlist.

We also show that human screening involves discretion and processing constraints, so applicants' outcomes may depend on the reviewer assigned to them. An AI screening tool reduces some of these constraints, such as order or workload effects, but it shifts our focus from heterogeneous treatment to standardization. Standardization does not ensure equal treatment. First, the model could also be biased. Therefore, the focus should shift first to

studying the model bias and how to correct for it.[40] Second even when bias-free standardization could create systematic disadvantages, particularly if employers converge on similar screening rules or adopt the same model.

Moreover, a common AI screener could miss useful information. In particular, human reviewers may be better at incorporating employers' needs beyond the objective criteria stated. However, in our setting, the narrowing of the hiring ethnic gap does not come with a reduction in overall hiring or satisfaction and match stability. At comparable screening capacity, the AI screener actually selects applicants with more hospitality experience and motivation, and applicants with immigrant-sounding names, in particular, have fewer employment gaps. These changes in observed characteristics, along with the finding that the reduction in the hiring ethnic gap comes from the matched-size set and persists when we reweight postings by shortlist size, suggest a role for applicant characteristics and job fit in narrowing the hiring ethnic gap. Instead, we find no evidence of improved ranking: immigrant rankings do not improve, and when we control for rank, and AI scores the estimated narrowing of the gap does not shrink.

At the same time, employers hire beyond the highest AI-ranked applicants. While hiring from the expansion set does not seem to drive the reduction in the ethnic hiring gap, our decomposition suggests it sustains the age penalty. These patterns suggest that the AI ranking does not fully capture employers' needs. We do not know what would have happened with size fixed to overall hiring, since employers might have substituted with other candidates. However, the hiring from the expansion set suggests a complementarity between algorithmic screening and human judgment at the final selection stage. This might help to sustain overall hiring i.e., finding suitable candidates, but also to sustain disparities attenuated at screening.

In addition, one possible response is to allow employers to personalize the AI. However, how to design this remains an open question, particularly which preferences should be elicited, which should be learned from observed choices, and which should be constrained. One approach is to use employers' stated weights on evaluation criteria, making personalization explicit and potentially auditable. Another could be to let the AI learn from a specific employer's past decisions, which brings us back to the fact that a model trained to reproduce those choices may reproduce barriers introduced at screening and hiring. Their consequences

[40] This is what we are currently doing with the Job matching Platform. First, a correspondence audit uses synthetic applications, randomized in their attributes and name, to study the bias and to inform the second experimental part. We correct the AI score ex post; previous research has found that changing the prompt could lead to overreaction Elass et al. (2026), and we randomize the score correction at the posting level.

for access to employment and matching are left for future research.

Finally, our paper shows that incorporating employer information into a screening rule is distinct from deciding who has authority to apply or override that rule. Our findings make salient the distinction between providing humans with AI-generated information and delegating decision-making authority to AI. In our main comparison, control-group reviewers already had access to AI scores, and delegating partial authority narrowed hiring ethnic gaps. In a separate comparison, the only difference in the regimes was whether human reviewers received AI information. Here, we find no detectable change in ethnic hiring disparities. In particular, we find that providing information may not be sufficient, and that information and discretion may operate simultaneously. We found that applicants with immigrant-sounding names still had lower shortlisting rates across all core bands, including when reviewers received the score, but the ethnic gap among high-scoring applicants narrowed when reviewers received information, consistent with some form of statistical discrimination. This pattern suggests that making information available may be insufficient when reviewers can still choose not to use it. Reviewers may also depart from AI recommendations because they have useful private information, disagree with the assessment, distrust the model, or face attention constraints. The distinction between providing humans with AI-generated information and delegating decision-making authority to AI speaks directly to the European Union AI Act and its regulation of the human oversight in algorithmic hiring. Our evidence suggests that a specific role assigned to humans matters and that the design should take into consideration not only screening but the whole pipeline.

10 Conclusion

This paper studies the delegation of applicant screening to AI by a labor market intermediary and its consequences for disparities in access to employment. Much of the debate about artificial intelligence and labor markets concerns which jobs and tasks AI replaces. However, we show that AI's effects on employment arise not only through changes in labor demand, but also through changes in how workers gain access to jobs, thereby affecting workers' outcomes even in occupations whose tasks are not necessarily affected by AI.

Using rich administrative data and three complementary research designs, we show that, in an intermediated labor market, under human screening, disparities arise at screening, persist into hiring, and vary across reviewers and employers. When instructed to apply the same criteria as humans and keeping screening capacity fixed, an AI screener selects different

candidates, and reduces disparities and dispersion across reviewers. Finally, moving from human reviewers to partial delegation of screening authority narrows the ethnic hiring gap without a detectable reduction in overall hiring. Employer satisfaction increases among respondents, while available post-hire measures show no clear deterioration. The pattern differs for age: even if screening does not penalize older applicants, they remain less likely to be hired. These contrasting patterns highlight why changes in screening must be evaluated through employers' subsequent choices and final hiring.

More broadly, our findings show that AI screening is a change in the allocation of opportunities, not merely a reduction in the cost of processing applications. They call for rethinking how disparities arise and operate as we move from heterogeneous human judgments to AI screening that applies a common screening rule. A common rule can reduce disparities arising from individual discretion while making any remaining penalties more systematic across employers. The consequences depend on both the screening authority and how employers respond to the candidates it selects. Providing AI information to human reviewers and delegating screening decisions are therefore distinct organizational choices. Downstream human judgment can preserve valuable matching opportunities, but it can also sustain disparities that screening attenuated. Understanding inequality and its changes under AI requires examining how evaluation criteria, decision authority, and employer choices jointly determine access to employment.

Some limitations remain. Our analysis does not causally identify the separate effects of shortlist size and applicant characteristics and job fit. Even though, the mechanism exercises help to describe where the changes in disparities come from, we do not know how employers would have responded to alternative shortlists. In addition, access to AI information and delegation of screening authority were also not independently randomized within a common population limiting what we can conclude about their relative effects and interaction. Finally, important measures of employer satisfaction and separation rates can be improved in their quality. Our results also point to future research and policy work. Firms and regulators can evaluate screening systems by examining both the rules governing access and the human decisions retained throughout hiring. Future work could study whether knowing AI selected the candidates changes employers' trust and subsequent choices. With the advancement of AI development, it becomes important to study the consequences on efficiency and disparities of personalizing screening to employers' stated priorities. One promising avenue combines systematic audits of AI evaluations with experiments of targeted corrections. Experiments that vary shortlist size, applicant characteristics and job fit, and

decision authority independently could identify designs that broaden access while preserving valuable matching opportunities.

References

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo, "Artificial intelligence and jobs: Evidence from online vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340.

Aggarwal, Varun, Kay Kobak, and John Howarter, "Using Large Language Models to Support High-Volume Application Review for an Undergraduate Research Program," 2026. Work-in-progress paper.

Allasino, Enrico, Emilio Reyneri, Alessandra Venturini, and Giovanna Zincone, "Labour market discrimination against migrant workers in Italy," International Migration Papers 67, International Labour Office, Geneva 2004.

Angelova, Victoria, Will Dobbie, and Crystal S Yang, "Algorithmic recommendations and human discretion," *Review of Economic Studies*, 2026, *93* (4), 2250–2283.

Arnold, David, Will Dobbie, and Peter Hull, "Measuring Racial Discrimination in Bail Decisions," *American Economic Review*, 2022, *112* (9), 2992–3038.

_ , _ , and _ , "Measuring Racial Discrimination in Bail Decisions," *American Economic Review*, September 2022, *112* (9), 2992–3038.

Arrow, Kenneth J., "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, Princeton University Press, 1973, pp. 3–33.

Awuah, Kobbina, Urša Krenk, and David Yanagizawa-Drott, "Augment or Automate? An Early Field Experiment with Generative AI in Hiring," Technical Report, University of Zurich 2026. Working paper, April 7.

Barbanchon, Thomas Le, Lena Hensvik, and Roland Rathelot, "How Can AI Improve Search and Matching? Evidence from 59 Million Personalized Job Recommendations," Technical Report, Working Paper 2023.

Baron, E. Jason, Joseph J. Doyle, Natalia Emanuel, Peter Hull, and Joseph Ryan, "Discrimination in Multiphase Systems: Evidence from Child Protection," *Quarterly Journal of Economics*, 2024, *139* (3), 1611–1664.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka, "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, 2016, *106* (6), 1437–1475.

Becker, Gary S., *The Economics of Discrimination*, University of Chicago Press, 1957.

Behaghel, Luc, Bruno Crépon, and Thomas Le Barbanchon, "Unintended Effects of Anonymous Résumés," *American Economic Journal: Applied Economics*, 2015, *7* (3), 1–27.

Ben-Michael, Eli, D. James Greiner, Melody Huang, Kosuke Imai, Zhichao Jiang, and Sooahn Shin, "Does AI Help Humans Make Better Decisions? A Statistical Evaluation Framework for Experimental and Observational Studies," Technical Report, Working Paper 2024.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 2004, *94* (4), 991–1013.

Bó, Ernesto Dal, Frederico Finan, Nicholas Y. Li, and Laura Schechter, "Information Technology and Government Decentralization: Experimental Evidence From Paraguay," *Econometrica*, 2021, *89* (2), 677–701.

Bohren, J. Aislinn, Peter Hull, and Alex Imas, "Systemic Discrimination: Theory and Measurement," *Quarterly Journal of Economics*, 2025, *140* (3), 1743–1799.

Brynjolfsson, Erik, Bharat Chandar, Ruyu Chen et al., "Canaries in the coal mine?: Six facts about the recent employment effects of artificial intelligence," 2025.

_ , Danielle Li, and Lindsey Raymond, "Generative AI at work," *The Quarterly Journal of Economics*, 2025, *140* (2), 889–942.

Buolamwini, Joy and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in "Conference on Fairness, Accountability, and Transparency" PMLR 2018, pp. 77–91.

Bursztyn, Leonardo, Thomas Chaney, Tarek A. Hassan, and Aakaash Rao, "The Immigrant Next Door," *American Economic Review*, 2024, *114* (2), 348–384.

Busetta, Giovanni, Maria Gabriella Campolo, and Demetrio Panarello, "Immigrants and Italian labor market: statistical or taste-based discrimination?," *Genus*, 2018, *74*, 4.

Dargnies, Marie-Pierre, Rustamdjan Hakimov, and Dorothea Kübler, "Aversion to Hiring Algorithms: Transparency, Gender Profiling, and Self-Confidence," Technical Report, WZB Discussion Paper 2022.

Elass, Kenza, Germain Gauthier, Debora Nozza, and Paola Profeta, "Automating Inequality? Gender Bias in AI-mediated Labor Markets," 2026. Unpublished manuscript.

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

Fernando, A. Nilesh, Niharika Singh, and Gabriel Tourek, "Hiring Frictions and the Promise of Online Job Portals: Evidence from India," *American Economic Review: Insights*, 2023, *5* (4), 546–562.

FIPE, "Ristorazione 2023: Rapporto annuale," Technical Report, Federazione Italiana Pubblici Esercizi 2023. Employment by nationality: annual average for 2022, based on INPS data.

Goldin, Claudia and Cecilia Rouse, "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review*, 2000, *90* (4), 715–741.

Hampole, Menaka, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller, "Artificial intelligence and the labor market," Technical Report, National Bureau of Economic Research 2025.

Harrington, Emma and Hannah Shaffer, "Learning About Police Bias: Prosecutors and Police Before and After Body-Worn Cameras," 2026. Working paper.

Hoffman, Mitchell, Lisa B. Kahn, and Danielle Li, "Discretion in Hiring," *Quarterly Journal of Economics*, 2018, *133* (2), 765–800.

Horton, John J., "The Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment," *Journal of Labor Economics*, 2017, *35* (2), 345–385.

ISTAT, "Employment and Labor Force Statistics," 2022. Italian National Institute of Statistics.

ISTAT, "Imprese e ICT: Anno 2025," Technical Report, Istituto Nazionale di Statistica 2025.

_ , "Settore I: Attività dei servizi di alloggio e di ristorazione," Scheda settoriale, Istituto Nazionale di Statistica 2025. Dati sulla struttura delle imprese e sull'occupazione riferiti al 2022.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human Decisions and Machine Predictions," *Quarterly Journal of Economics*, 2018, *133* (1), 237–293.

Kline, Patrick, Evan K. Rose, and Christopher R. Walters, "Systemic Discrimination Among Large U.S. Employers," *Quarterly Journal of Economics*, 2022, *137* (4), 1963–2036.

_ **,** _ **, and** _ , "A Discrimination Report Card," *American Economic Review*, 2024, *114* (8), 2472–2525.

_ **, Raffaele Saggio, and Mikkel Sølvsten**, "Leave-out estimation of variance components," *Econometrica*, 2020, *88* (5), 1859–1898.

Knowles, John, Nicola Persico, and Petra Todd, "Racial bias in motor vehicle searches: Theory and evidence," *Journal of political economy*, 2001, *109* (1), 203–229.

Krueger, Alan B. and Lawrence H. Summers, "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica*, 1988, *56* (2), 259–293.

Leibbrandt, Andreas, Mallory Avery, Edwina Ip, and Joseph Vecci, "A Brave New World of Hiring: A Natural Field Experiment on How Asynchronous Interviews and AI Assessment Reshape Recruitment," 2026.

Li, Danielle, Lindsey Raymond, and Peter Bergman, "Hiring as Exploration," *Review of Economic Studies*, 2026, *93* (2), 1200–1240.

Lippens, Louis, "Computer Says 'No': Exploring Systemic Bias in ChatGPT Using an Audit Approach," *Computers in Human Behavior: Artificial Humans*, 2024, *2* (1).

Ludwig, Jens and Sendhil Mullainathan, "Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System," *Journal of Economic Perspectives*, 2021, *35* (4), 71–96.

Mehmood, Sultan, Christoph Goessmann, and Elliott Ash, "Courts of tomorrow: Evidence from a nationwide rollout of generative AI," 2026.

Miller, Jim, "Reviewing 1500 Resumes in 6 Hours with AI-Assisted Application Review in Ashby," Ashby September 2024.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Peter Barnes, Lucy Vasserman, Ben Hutchinson, Elena Barnes, Maarten Sap, Webster Lebrecht, and Joy Buolamwini, "Model Cards for Model Reporting," *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019, pp. 220–229.

Mocanu, Tatiana, "Designing Gender Equity: Evidence from Hiring Practices," IZA Discussion Paper 17480, IZA Institute of Labor Economics November 2024.

Neumark, David, "Experimental Research on Labor Market Discrimination," *Journal of Economic Literature*, 2018, *56* (3), 799–866.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science*, 2019, *366* (6464), 447–453.

Phelps, Edmund S., "The Statistical Theory of Racism and Sexism," *The American Economic Review*, 1972, *62* (4), 659–661.

Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H. Midtbøen, "Meta-analysis of Field Experiments Shows No Change in Racial Discrimination in Hiring over Time," *Proceedings of the National Academy of Sciences*, 2017, *114* (41), 10870–10875.

SHRM, "2025 Talent Trends," 2025.

Teeselink, Bouke Klein, "Generative AI and Labor Market Outcomes: Evidence from the United Kingdom," Working Paper, King's College London 2025. Version dated December 21, 2025.

Tomlinson, Kiran, Sonia Jaffe, Will Wang, Scott Counts, and Siddharth Suri, "Working with AI: Measuring the Applicability of Generative AI to Occupations," 2025.

Unioncamere and Ministero del Lavoro e delle Politiche Sociali, "La domanda di professioni e di formazione delle imprese italiane nel 2025," Sistema Informativo Excelsior, Unioncamere, Roma 2025.

Voth, Joachim and Guo Xu, "Discretion and Destruction: Promotions, Performance, and Patronage in the Royal Navy," Working paper, University of Zurich and University of California, Berkeley 2022. February 24, 2022 version.

Appendix

Contents

A Additional Tables and Figures **54**

B Data Appendix **54**

B.1 Control Variables 55

B.2 Immigrant sounding name indicator 57

C Construction of WhatsApp measures **58**

C.1 Dimensions and scoring anchors 58

C.2 Model and inference procedure 59

D The production AI rubric **61**

E Descriptives **61**

F Human-screening robustness **68**

F.1 Information and processing constraints 72

F.2 Predictive importance 78

F.3 Flexible adjustment 79

F.4 Post-hire comparisons in the Human-only sample 79

F.5 Missing information and the name gap 80

G Organizational heterogeneity: estimation and interpretation **81**

G.1 Dispersion and empirical Bayes 81

G.2 Reviewer specific weights on applicant characteristics 84

G.3 Joint reviewer and employer slopes 90

G.4 Crossed pair-gap decomposition 94

G.5 Leave-posting-out variance correction 95

H Flexible same-applicant comparisons **96**

H.1 Agreement at equal screening capacity 96

H.2 Comparing conditional selection associations 97

H.3 Flexible adjustment 98

H.4 Reviewer gaps on a common vacancy distribution 100
H.5 Workload under the two screening rules 101

I AI-Hybrid versus Human + Information 102
I.1 Mechanisms: Ranking, shortlist size, applicant characteristics, and employer decisions 109

J Other deployment comparisons 126
J.1 Human + Information versus Human only 126
J.2 Fully AI versus AI-Hybrid 134

K A Framework for Screening Signals and Employer Response 143
K.1 Delegation and information 143
K.2 Applicant fit and screening assessments 144
K.3 Shortlist and employer consideration 145
K.4 Employer assessment and choice 145
K.5 Conditional implications 145
K.6 Mapping to the empirical analysis 147

A Additional Tables and Figures

B Data Appendix

Table B1: Data structure and coverage.

Data layer	Variables	Availability
Workers	Demographics and location Job and contract preferences Education and language proficiency Employment history and experience CV content, quality, and organization Immigrant sounding name indicator	Full period
Job postings	Occupation and job title Salary and benefits Contract type, hours, and shifts Experience and language requirements Location and commuting requirements	Full period
Applications	Application and decision dates Shortlisting, interview, offer, and hiring outcomes WhatsApp assessment transcript Rejection reason Measures of match between applicants and jobs Presentation rank	Full period
Screening	Responsible CS reviewer Human screening decision AI composite score	Full period
Employer response	Candidate profile opened Interview Offer Hire	Full period
Post hire	Worker reapplication Employer reposting of the vacancy Composite separation proxy Employer turnover survey	Hired applications only Follow up varies by hire date; survey response is selective
Firms and management	Sector, legal form, and registered location Employment, revenues, and assets Management and governance measures Named manager or director characteristics Candidate firm outcomes	AIDA linked records 57.9% of eligible employers 62.1% of postings

B.1 Control Variables

Our main analysis includes applicant, CV characteristics, worker preferences, and measures of applicant–vacancy match. Missing-value indicators are included where appropriate to retain observations without treating unavailable information as an unfavorable observed characteristic.

Applicant and CV characteristics. The applicant-level controls are constructed from platform registration records and information extracted from uploaded CVs.

- *Demographics and human capital.* Age at application, gender, presence of a platform profile picture, Italian and English proficiency at B2 or above, number of languages, years of education, foreign education, hospitality-specific education, professional certifications, and current student status. Age is calculated from the platform birthdate, with CV-derived age used when unavailable. Gender is obtained from platform records, supplemented by NamSor when missing.

- *Employment history.* Current unemployment status, years of hospitality and non-hospitality experience, years since the last hospitality job, employment gaps, whether the applicant's career has been entirely in hospitality, and managerial or international work experience. Employment histories are reconstructed from the occupations and dates reported in CVs.

- *CV quality and presentation.* CV structure score, presence of a photograph, Europass format, consistency of CV language, document length, and the fraction of employment spells with recorded dates.

- *Presentation of job-relevant information.* Whether the occupation required by the vacancy appears on the first page of the CV, whether work experience precedes education, years of experience in the required occupation, and indicators for missing experience, education, or language sections.

- *Preferences and mobility.* Preferences for full-time and part-time employment, willingness to relocate permanently or seasonally, unwillingness to relocate, and availability of a car or public transportation.

- *Application order.* The applicant's position in the sequence of applications to the same vacancy, normalized as $(r_{ij} - 1)/(N_j - 1)$, where r_{ij} is the applicant's rank in

submission order and N_j is the number of applications to vacancy j. The measure equals zero for vacancies with a single application. The common specification also includes the accompanying order and competition controls.

Applicant–vacancy match. We combine applicant characteristics and stated preferences with vacancy requirements to construct measures of compatibility. These measures are independent of the production AI screening score.

- *Occupation match.* Indicator for whether the applicant's occupation group matches the occupation required by the vacancy.
- *Experience match.* Indicator for whether the applicant meets the vacancy's stated experience requirement.
- *Minimum hospitality experience match.* Indicator for whether the applicant's years of hospitality experience meet the minimum experience required by the vacancy.
- *Occupation–experience match.* Indicator for whether the applicant both matches the required occupation group and meets the minimum experience requirement.
- *Distance match.* Indicator for whether the applicant resides within the maximum commuting distance specified by the vacancy.
- *Required-language match.* Indicator for whether the applicant meets the vacancy's language requirements.
- *Language-level match.* Indicator for whether the applicant meets the required language proficiency levels.
- *Certification match.* Indicator for whether the applicant holds the certifications required by the vacancy.
- *Sector match.* Indicator for whether the vacancy's sector is among the applicant's preferred sectors.
- *Working-hours match.* Indicator for whether the offered full-time or part-time position matches the applicant's preferences.
- *Salary match.* Indicator for whether the offered salary is within the applicant's preferred salary tier.

- *Shift match.* Indicator for whether at least one of the applicant's preferred shifts overlaps with the vacancy's service slots.

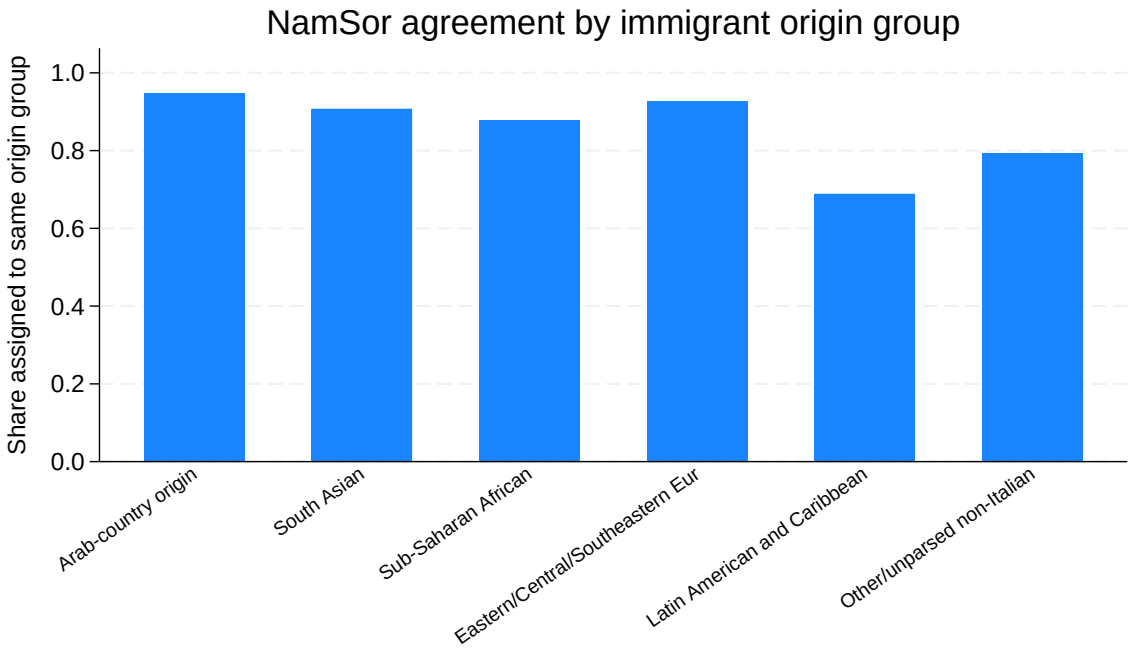
Our preferred specification also includes five match indicators separately: occupation, experience, required language, distance, and sector. As an alternative, we construct a *match composite* equal to the unweighted average of these five indicators. The index ranges from zero to one and averages the available components for each applicant–vacancy pair. Missing components are excluded from the denominator, and the index is missing only when all five components are missing.

B.2 Immigrant sounding name indicator

We use NamSor's full-name-country service to classify names as immigrant-sounding. The indicator equals zero when either primary or alternative country prediction is Italy, 0 otherwise. If there are not classifiable names they are missing.

We compare this indicator with the nationality stated in the CVs, when available. Among the 55,116 CVs that report nationality, the two measures agree in 86.4 percent of cases. NamSor classifies 98.0 percent of applicants with non-Italian CV nationality as having immigrant sounding names and 79.1 percent of applicants with Italian CV nationality as having Italian sounding names. When we use origin groups the agreement is 80.4 percent. Figure B1. Names and nationality need not coincide. However, as a robustness check for the main tables, we use an alternative indicator based on CV nationality when observed and the NamSor classification when nationality is missing.

Figure B1: Name classification and CV-stated nationality



C Construction of WhatsApp measures

In this section we document our procedure that converted WhatsApp pre-assessment transcripts into controls for our analysis. We have 131,860 WhatsApp pre-assessment. Importantly, we did not use for the classification any other information aside the WhatsApp text.

C.1 Dimensions and scoring anchors

We classify each transcript along ten prespecified dimensions: (i) availability to start work; (ii) shift availability; (iii) Italian proficiency; (iv) English proficiency; (v) relevant work experience; (vi) distance, transportation, and commute feasibility; (vii) soft skills; (viii) role-specific technical skills; (ix) motivation for the role; and (x) reply quality.

Each transcript is processed twice using separate prompts, and we create categorical and numerical variables generated independently. We check the internal consistency among the two to validate our procedure. For the categorical sentiment analysis for the first nine dimensions, the labels allowed are *positive*, *mixed*, *negative*, and *neutral*; neutral means that the transcript contains no relevant information. Instead, reply quality uses *detailed*, *adequate*,

vague, and *minimal*. The numerical pass returns a score from 0 to 10 and supporting evidence.

Table C1 reports the prespecified numerical anchors. Score 0 indicates missing relevant information. The higher the score the better the fit.

Table C1: Numerical anchors for transcript-based assessment measures

Dimension	Numerical anchors
Availability and shift availability	0: not discussed; 1: unavailable; 2–4: partially available; 5–7: available with some constraint; 8–10: fully available.
Italian and English	0: not discussed; 1: no spoken at all; 2–4: basic; 5–7: intermediate; 8–10: fluent.
Relevant work experience	0: not discussed; 1: no relevant experience; 2–5: partial experience; 6–8: one to two years; 9–10: at least three years.
Distance and commute	0: not discussed; 1: incompatible; 2–4: logistically difficult; 5–7: feasible; 8–10: close and/or independently mobile.
Soft skills	0: not discussed; 1: clearly poor; 2–4: weak or uncertain; 5–7: adequate; 8–10: strong.
Role-specific technical skills	0: not discussed; 1: required skills absent; 2–4: limited ; 5–7: some required skills; 8–10: strong skills present.
Motivation	0: not discussed; 1: absent; 2–4: low or uncertain; 5–7: adequate; 8–10: strong.
Reply quality	0: no substantive reply; 1: minimal; 2–4: vague; 5–7: adequate; 8–10: detailed.

C.2 Model and inference procedure

Our classification proceeds as follows: a prompt containing the task, allowed labels, scoring rules, and required output calls `gemma-4-12B-it` model with temperature zero. The output is a JSON file containing the label or the score and a short explanation. Using 33 transcripts we assess the sensitivity to the prompt structure: (i) we process these transcripts both with a single prompt covering all the ten dimensions and with separate prompts, one for each dimension; (ii) we check categorical labels and numerical scores. Table C2 reports agreement between categorical labels and the mean absolute difference between numerical scores under the two prompt structures.

Categorical agreement ranges from 54.5 percent for soft skills to 100 percent for Italian proficiency. We do not observe any direct reversal between positive and negative labels for

any of the dimensions. There could be still some differences in numerical scores. For example, for the dimension of Italian proficiency, the labels agree in all 33 cases, but the mean absolute score difference is 2.21 points on the 0–10 scale. Our checks are good to document sensitivity to the prompt structure and consistency between categorical and numerical values. However, they do not replace an assessment done by humans.

Table C2: Sensitivity of WhatsApp measures to prompt structure

Dimension	Identical labels (out of 33)	Agreement (percent)	Mean absolute score difference
Availability	31	93.9	0.24
Shift availability	30	90.9	0.33
Italian proficiency	33	100.0	2.21
English proficiency	30	90.9	0.21
Relevant work experience	28	84.8	0.88
Distance and commute	29	87.9	1.30
Soft skills	18	54.5	0.76
Technical skills	28	84.8	0.85
Motivation	27	81.8	1.12
Reply quality	25	75.8	0.94

Notes: Sample 33 transcripts. Column (1) reports the number of transcript in which there is agreement between the labels given by the combined prompt and separate prompts. Column (2) reports the percentage. Column (3) reports the difference in numerical scores.

D The production AI rubric

Table D1: The nine production scoring criteria

Area (weight)	Criterion	Illustrative anchors
Job match (40%)	Experience in role	10: 3+ years; 8: 1–2 years; 5: related role; 1: none.
	Language	10: fluent; 7: intermediate; 4: basic; 1: none.
	Availability	10: aligned; 7: minor difference; 4: partial; 1: incompatible.
Soft skills and attitude (30%)	Professionalism	10: excellent; 5–6: concise but respectful; 1–2: rude.
	Motivation	10: strong; 5–6: not explicit, no disinterest; 1–2: hostile.
	Reliability	10: proven; 7: consistent; 4: unclear; 1: absent.
Technical and logistical fit (30%)	Practical skills	10: all required; 7: some; 4: learnable; 1: none.
	Job stability	10: at least 12 months per job; 7: 6–12; 4: short jobs; 1: major gaps.
	Logistics and distance	10: within radius; 7: borderline; 4: difficult; 1: incompatible.

E Descriptives

Figure E1: Wage variation between and within occupations. Differences between occupational categories account for 44 percent of salary dispersion across full time vacancies; variation within occupations accounts for 56 percent.



Table E1: Worker characteristics by name group

	All Mean	Italian-sounding name Mean	Immigrant-sounding name Mean	Difference (Imm. minus Ital.)
Demographics and human capital				
Female	0.381	0.382	0.379	-0.003
Age at application (years)	32.994	35.112	30.267	-4.845***
Italian proficiency B2 or above/native	0.891	1.000	0.693	-0.307***
English proficiency B2 or above	0.578	0.550	0.614	0.063***
University degree or above	0.253	0.209	0.313	0.104***
Foreign education	0.066	0.009	0.135	0.125***
Experience				
Total work experience (years)	8.506	10.574	5.969	-4.605***
Hospitality experience (years)	4.808	6.068	3.262	-2.806***
Distinct hospitality employers	3.104	3.747	2.316	-1.430***
Any employment gap	0.554	0.618	0.475	-0.143***
Management experience	0.282	0.346	0.202	-0.145***
CV quality and completeness				
CV structure score (0–10)	5.321	5.431	5.186	-0.245***
CV sections detected	3.963	4.246	3.620	-0.626***
CV length is one or two pages	0.517	0.563	0.461	-0.102***
Preferences and logistics				
Owns or uses a car	0.569	0.701	0.355	-0.345***
Prefers part-time work	0.401	0.371	0.451	0.080***
Workers	58,877	32,468	26,409	

Notes: This table presents descriptive statistics for workers in the Human only sample (See Section 4). Column 1 presents statistics for all workers; Columns 2 and 3 present statistics separately for workers with Italian-sounding names and workers with Immigrant sounding names. Column 4 reports the standardized differences, with stars reporting reports p-values from tests of equality of means. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table E2: Applicant–vacancy match by name group

	All Mean	Italian-sounding name Mean	Immigrant-sounding name Mean	Difference (Imm. minus Ital.)
Panel A: Candidate–job match indicators				
Candidate–job match composite	0.465	0.490	0.437	-0.054***
Occupation match	0.485	0.523	0.440	-0.083***
Meets minimum-experience requirement	0.578	0.628	0.521	-0.107***
Meets language requirement	0.641	0.633	0.653	0.020**
Meets distance requirement	0.786	0.794	0.777	-0.017***
Applications	121,897	65,170	56,727	
Panel B: Components constructed from the AI rubric				
Experience component	3.427	3.682	3.135	-0.547***
Practical-skills component	4.627	4.975	4.227	-0.748***
Language-match component	4.498	4.430	4.582	0.152**
Distance component	8.253	8.323	8.172	-0.151***
Job-stability component	4.564	4.139	5.076	0.937***
Applications	121,897	65,170	56,727	

Notes: This table presents descriptive statistics for applications in the Human only sample (See Section 4). Column 1 presents statistics for all applications; Columns 2 and 3 present statistics separately for application from workers with Italian-sounding names and workers with Immigrant sounding names. Column 4 reports the standardized differences, with stars reporting reports p-values from tests of equality of means. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table E3: Vacancy sorting: jobs and contracts

	All Mean	Italian-sounding name Mean	Immigrant-sounding name Mean	Difference (Imm. minus Ital.)
Panel A: Occupation and skill content				
Skill tier: senior/lead/management role	0.109	0.145	0.067	-0.079***
Skill tier: low-skill/unskilled role	0.122	0.111	0.135	0.023***
Panel B: Contract and work arrangement				
Contract: permanent (tempo indeterminato)	0.089	0.100	0.076	-0.024***
Contract: temp with permanent prospect	0.639	0.623	0.657	0.033***
Contract: short-term (6 months or less)	0.225	0.233	0.216	-0.016***
Contract: on-call	0.006	0.006	0.007	0.001***
Full time (40h/week)	0.872	0.873	0.872	-0.001
Part time	0.124	0.123	0.124	0.001
Seasonal position	0.238	0.243	0.232	-0.012***
Panel C: Requirements and compensation				
Min. experience requested (years)	1.732	1.841	1.608	-0.233***
Monthly net-equivalent salary	1686.577	1745.200	1619.229	-125.971***
Position located outside Italy	0.131	0.156	0.102	-0.054***
Applications	121,897	65,170	56,727	

Notes: This table presents descriptive statistics for applications in the Human only sample (See Section 4). Column 1 presents statistics for all applications; Columns 2 and 3 present statistics separately for application from workers with Italian-sounding names and workers with Immigrant sounding names. Column 4 reports the standardized differences, with stars reporting reports p-values from tests of equality of means. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table E4: Vacancy sorting: benefits and establishments

	All Mean	Italian-sounding name Mean	Immigrant-sounding name Mean	Difference (Imm. minus Ital.)
Benefit: meals included	0.806	0.803	0.809	0.006**
Benefit: housing provided	0.429	0.458	0.395	-0.063***
Benefit: tips	0.104	0.105	0.103	-0.002
Benefit: performance bonus	0.083	0.088	0.077	-0.012***
Restaurant: hotel/resort	0.179	0.187	0.170	-0.017***
Restaurant: formal/starred	0.094	0.095	0.092	-0.003*
Restaurant: pizzeria	0.178	0.171	0.187	0.017***
Restaurant: bar/cocktail/pub	0.177	0.186	0.167	-0.019***
Restaurant: fast food/tavola calda	0.075	0.080	0.069	-0.011***
Restaurant: trattoria/osteria/agriturismo	0.085	0.080	0.090	0.009***
covers per service (midpoint)	104.821	105.486	104.060	-1.426***
Average price per cover (EUR)	29.958	29.913	30.009	0.096
Applications	121,897	65,170	56,727	

Notes: This table presents descriptive statistics for applications in the Human only sample (See Section 4). Column 1 presents statistics for all applications; Columns 2 and 3 present statistics separately for application from workers with Italian-sounding names and workers with Immigrant sounding names. Column 4 reports the standardized differences, with stars reporting reports p-values from tests of equality of means. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table E5: Missing worker and CV characteristics

	All Mean	Italian-sounding name Mean	Immigrant-sounding name Mean	Difference (Imm. minus Ital.)
Panel A: Demographics and human capital				
Missing: Female	0.570	0.534	0.613	0.078***
Missing: Age at application (years)	0.135	0.117	0.157	0.039***
Missing: Italian proficiency B2 or above/native	0.147	0.000	0.328	0.328***
Missing: English proficiency B2 or above	0.400	0.383	0.421	0.038***
Missing: University degree or above	0.161	0.123	0.208	0.084***
Missing: Foreign education	0.022	0.023	0.020	-0.002*
Panel B: Experience				
Missing: Total work experience (years)	0.022	0.023	0.020	-0.002*
Missing: Hospitality experience (years)	0.022	0.023	0.020	-0.002*
Missing: Distinct hospitality employers	0.022	0.023	0.020	-0.002*
Missing: Any employment gap	0.022	0.023	0.020	-0.002*
Missing: Management experience	0.076	0.071	0.082	0.011***
Panel C: CV quality and completeness				
Missing: CV structure score (0–10)	0.008	0.008	0.008	-0.001
Missing: CV completeness score	0.031	0.032	0.031	-0.001
Missing: Photo on CV	0.655	0.665	0.643	-0.022***
Missing: Number of CV pages	0.011	0.012	0.011	-0.001
Missing: CV sections detected	0.391	0.395	0.386	-0.009**
Missing: CV length is one or two pages	0.011	0.012	0.011	-0.001
Workers	58,877	32,468	26,409	

Notes: This table presents descriptive statistics of missing values for workers in the Human only sample (See Section 4). Column 1 presents statistics for all workers; Columns 2 and 3 present statistics separately for workers from workers with Italian-sounding names and workers with Immigrant sounding names. Column 4 reports the standardized differences, with stars reporting reports p-values from tests of equality of means. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

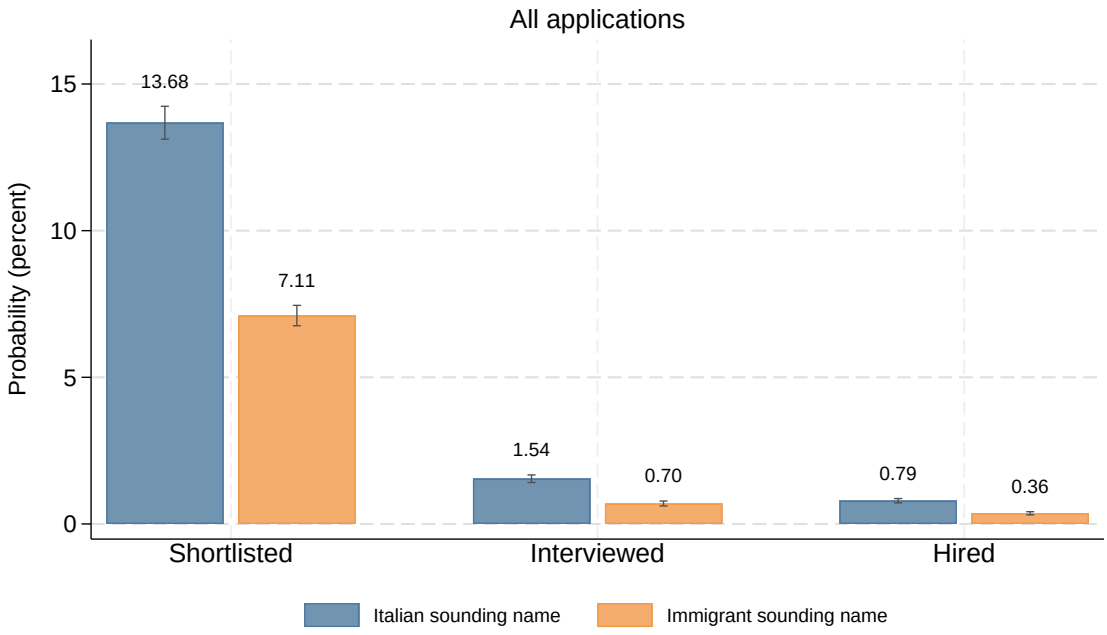
Table E6: Missing posting and match information

	All Mean	Italian-sounding name Mean	Immigrant-sounding name Mean	Difference (Imm. minus Ital.)
Panel A: Posting characteristics				
Missing: Min. experience requested (years)	0.004	0.004	0.005	0.002***
Missing: Monthly net-equivalent salary	0.000	0.000	0.000	0.000
Missing: Contract: permanent (tempo indeterminato)	0.007	0.006	0.007	0.001
Missing: Full time (40h/week)	0.004	0.004	0.003	-0.001*
Missing: Seasonal position	0.000	0.000	0.000	0.000
Missing: Skill tier: senior/lead/management role	0.002	0.003	0.001	-0.002***
Missing: covers per service (midpoint)	0.126	0.128	0.125	-0.003
Missing: Average price per cover (EUR)	0.227	0.231	0.221	-0.010***
Panel B: Candidate–job match				
Missing: Candidate–job match composite	0.000	0.000	0.000	0.000
Missing: Occupation match	0.008	0.009	0.008	-0.001
Missing: Meets minimum-experience requirement	0.032	0.031	0.033	0.002
Missing: Meets language requirement	0.611	0.595	0.630	0.034***
Missing: Meets distance requirement	0.028	0.027	0.030	0.004*
Missing: Match: job sector = worker preferred sector (0/1)	0.077	0.080	0.074	-0.006*
Missing: Match: position salary in worker's preferred tier (approx, 0/1)	0.076	0.078	0.073	-0.005*
Missing: Match: worker prefers the offered job type (0/1)	0.789	0.761	0.820	0.058***
Missing: Match: any worker shift pref overlaps position service slot (0/1)	0.797	0.778	0.820	0.042***
Missing: match_certificate	0.073	0.088	0.055	-0.033***
Applications	121,897	65,170	56,727	

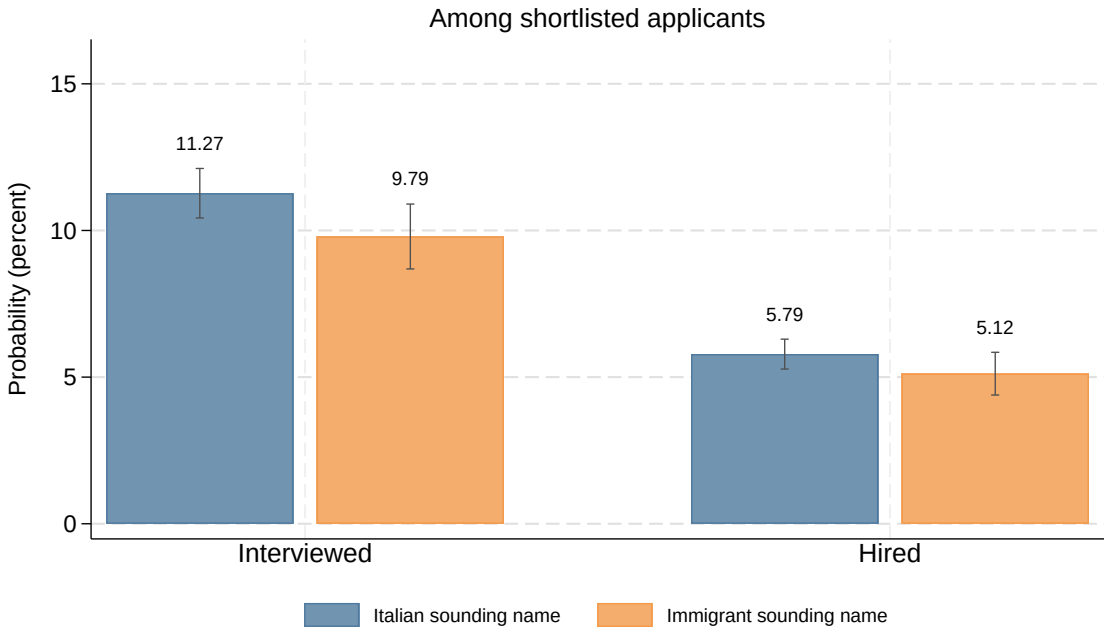
Notes: This table presents descriptive statistics for missing values of applications in the Human only sample (See Section 4). Column 1 presents statistics for all applications; Columns 2 and 3 present statistics separately for application from workers with Italian-sounding names and workers with Immigrant sounding names. Column 4 reports the standardized differences, with stars reporting reports p-values from tests of equality of means. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

F Human-screening robustness

Figure F1: Unadjusted outcomes through the human screening pipeline



(a) Outcomes in the full applicant pool



(b) Outcomes among shortlisted applicants

Notes: This figure reports unadjusted group means in outcomes, separately for applications with Italian sounding name (blue) and immigrant sounding names (orange). Panel (a) uses the 123,311 applications to 2,586 postings in the preferred human screening sample. Panel (b) restricts this sample to the 13,090 shortlisted applications. Horizontal bars show 95 % confidence intervals with clustered standard errors at the posting level.

Table F1: Interview and hiring gaps, unconditional and among shortlisted applicants

	Interviewed		Hired	
	(1)	(2)	(3)	(4)
Immigrant-sounding name	-0.005***	-0.005	-0.003***	-0.007
	(0.001)	(0.007)	(0.001)	(0.005)
Sample	All	Shortlisted	All	Shortlisted
Job-posting and Month-Year FE	✓	✓	✓	✓
Applicant and CV controls	✓	✓	✓	✓
Candidate–job match controls	✓	✓	✓	✓
WhatsApp assessment scores	✓	✓	✓	✓
Applications	123,311	12,778	123,311	12,778
Job postings	2,586	2,035	2,586	2,035
Mean Y (Italians)	0.015	0.111	0.008	0.057
Adjusted R^2	0.042	0.140	0.019	0.052

Notes. This table reports regression estimates of equation 1 using as dependent variable Interview in Columns (1) and (2), and Hired in Columns (3) and (4). Columns (1) and (3) use as sample all the applications, while Columns (2) and (4) restrict to shortlisted applications $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table F2: Age gradients in the Human-only sample

	Shortlisted	Interviewed	Hired
Above-median age	-0.0382^{***}	-0.0054^{***}	-0.0037^{***}
	(0.0024)	(0.0008)	(0.0006)
Age: ten additional years	-0.0289^{***}	-0.0036^{***}	-0.0021^{***}
	(0.0010)	(0.0003)	(0.0002)
Job-posting and Month-Year FE	✓	✓	✓
Applicant and CV controls	✓	✓	✓
Candidate–job match controls	✓	✓	✓
WhatsApp assessment scores	✓	✓	✓
Applications	123,311	123,311	123,311
Job postings	2,586	2,586	2,586
Mean Y (at/below-median age)	0.112	0.013	0.007
Adj. R^2: above-median age	0.153	0.041	0.018
Adj. R^2: continuous age	0.158	0.042	0.019

Notes. This table reports regression estimates of equation 1 using as dependent variables Shortlisted (Column (1)) , Interview (Column (2)), Hired (Column (3)). The sample uses all the applications. The first row reports coefficients for the indicator of above median age, the second row for additional ten years of age $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table F3: Robustness: combined CV-nationality and NamSor indicator

	Dep. var.: Shortlisted (0/1)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Non-Italian	-0.066***	-0.062***	-0.046***	-0.047***	-0.043***	-0.041***	-0.038***	-0.036***	-0.036***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Match composite								0.168***	
								(0.005)	
Job-posting and Month-Year FE	No	✓	✓	✓	✓	✓	✓	✓	✓
WhatsApp assessment scores	No	No	✓	✓	✓	✓	✓	✓	✓
Demographics	No	No	No	✓	✓	✓	✓	✓	✓
Experience	No	No	No	No	✓	✓	✓	✓	✓
CV quality + attention	No	No	No	No	No	✓	✓	✓	✓
Preferences	No	No	No	No	No	No	✓	✓	✓
Match dimensions + order	No	No	No	No	No	No	No	No	✓
Match composite + order	No	No	No	No	No	No	No	✓	No
Applications	123,326	123,311	123,311	123,311	123,311	123,311	123,311	123,311	123,311
Mean Y (Italian)	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135
Adj. R^2	0.011	0.068	0.116	0.127	0.134	0.137	0.139	0.150	0.158

Notes. This table reports regression estimates of equation 1 using as dependent variable an indicator for being shortlisted. The Non-Italian indicator uses CV stated nationality when observed and the NamSor classification otherwise. Columns reproduce the ones in Table 1. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table F4: Robustness: nonlinear age and WhatsApp controls

	Linear	+ Age squared	+ Score categories	+ Experience squares
Immigrant-sounding name	-0.037***	-0.039***	-0.038***	-0.039***
	(0.002)	(0.002)	(0.002)	(0.002)
Preferred controls	Yes	Yes	Yes	Yes
Posting and month-year FE	Yes	Yes	Yes	Yes
Age squared	No	Yes	Yes	Yes
WhatsApp scores	Linear	Linear	Categorical	Categorical
Experience squares	No	No	No	Yes
Control mean: Italian-sounding	0.137	0.137	0.137	0.137
Applications	123,311	123,311	123,311	123,311
Adjusted R-squared	0.158	0.163	0.164	0.164
Joint p-value: score nonlinearity	.	.	0.000	0.000

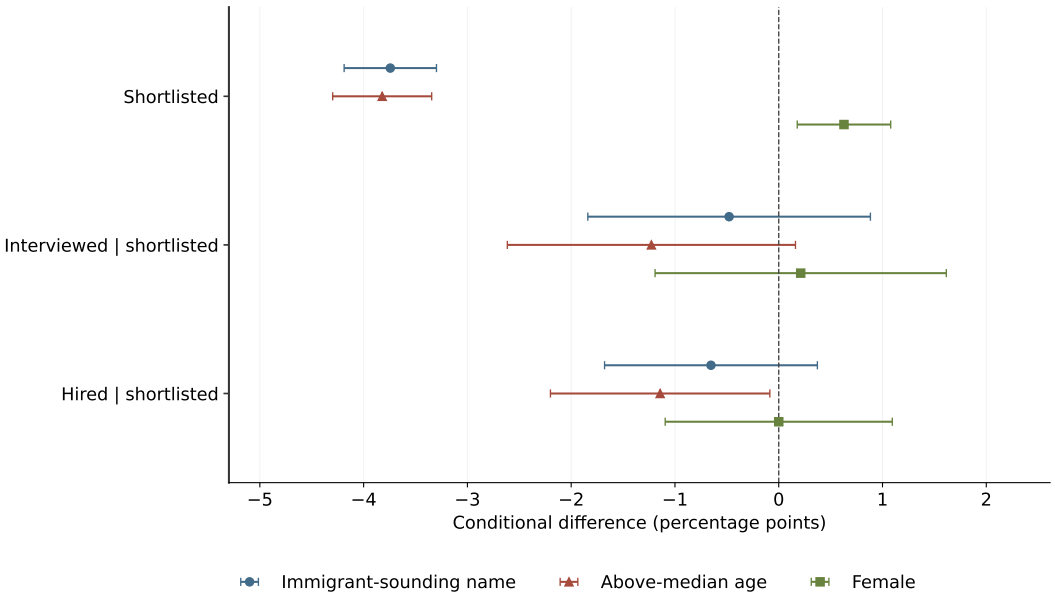
Notes. This table reports regression estimates of equation 1 using as dependent variable an indicator for being shortlisted. Columns add age squared, replace linear WhatsApp scores with categories, and add experience squares. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table F5: Conditional shortlisting gaps by name-origin group

	Arab-country	South Asian	Sub-Saharan African	Eastern/Central/ SE European	Latin American/ Caribbean	Other
Immigrant-sounding name	-0.058***	-0.086***	-0.065***	-0.021***	-0.012***	-0.028***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Job-posting FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Preferred controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Y (Italian-sounding)	0.137	0.137	0.137	0.137	0.137	0.137
Applications	78,685	75,734	75,263	72,438	75,140	75,074
Adjusted R-squared	0.166	0.169	0.167	0.169	0.168	0.168

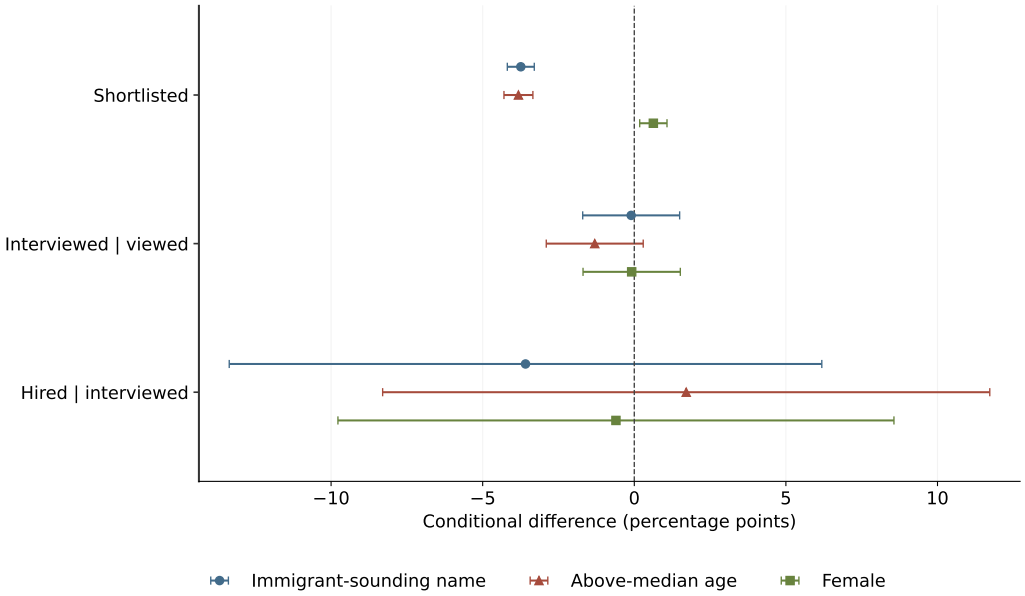
Notes. This table reports regression estimates of equation 1 using as dependent variable an indicator for being shortlisted. Each column compares one NamSor origin group with applicants with Italian sounding names. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure F2: Demographic gaps at screening and among shortlisted applicants



Notes. This figure show regression estimates of equation 1. Shortlisting estimates use all applications, interview and hiring estimates use only shortlisted applications. Horizontal bars show 95 % confidence intervals with clustered standard errors at the posting level.

Figure F3: Demographic gaps conditional on prior stages



Notes. This figure show regression estimates of equation 1.Shortlisting estimates use all applications, interview uses only opened profiles, and hiring estimates use only interviewed applicants. Horizontal bars show 95 % confidence intervals with clustered standard errors at the posting level.

F.1 Information and processing constraints

Differences across reviewers may reflect the hiring contexts in which they operate, as well as differences in information, working conditions, and judgments.

Vacancy characteristics. We first examine how ethnic shortlisting gaps vary across observable vacancy characteristics (Appendix Figure F4). The conditional ethnic shortlisting gap is smaller in customer-facing roles than in other roles (3.14 versus 4.61 percentage points). Separate PPML specifications confirm a smaller relative gap in these roles. When interactions with occupation, geographic area, and contract type enter jointly, the estimated difference retains the same sign but is no longer statistically significant. The point estimates therefore run against the prediction of anticipated costumer preferences (Appendix Table F6). Relative gaps are also larger in the South and islands than in northern Italy, after accounting for observed applicant and match characteristics and posting fixed effects. This pattern is consistent with reviewers anticipating differences in employer preferences across local labor markets. Permanent positions exhibit larger relative gaps than non-permanent positions, even though the difference in percentage-point gaps is not statistically significant (Appendix Table F6).

Missing applicant information. We then show that the ethnic screening gap is larger when productive information about an applicant is missing. When the experience section is missing, the name gap is 5.1 percentage points compared with 3.7 percentage points when it is present in the CV. Adjusting for multiple testing and entering all five missing-section interactions jointly do not change the significance of the difference (Appendix Tables F13 and F14). The pattern suggests that reviewers rely more on perceived origin when they lack direct information about relevant qualifications.

Reviewer workload. Even when information is available, reviewers may face constraints on their ability to process it. A heavier workload could lead to a higher use of group-based assessments, possibly widening ethnic screening gaps. Workload differences could also help explain why some reviewers show larger gaps than others. We examine both possibilities using applications arriving during the preceding seven days at the reviewer's other postings as a measure of workload, expressed as the standardized logarithm of one plus this count.

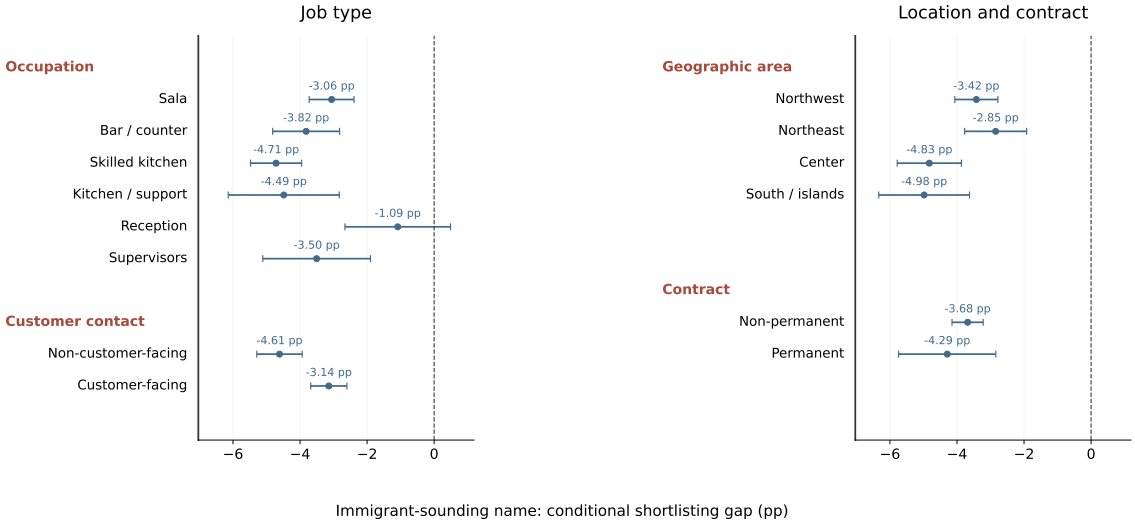
A one-standard-deviation increase in workload is associated with a 0.46-percentage-point larger ethnic gap when reviewer fixed effects are included. When we allow for reviewer-specific name gaps the estimate is 0.35 points and imprecise (Appendix Table F9). Therefore, we do not have enough information. Responses nevertheless differ across reviewers: we reject equality of workload gradients even when we allow the workload slope for Italian-sounding applicants to vary by reviewer ($p = 0.0155$; Appendix Figure G8). Workload may therefore affect screening disparities distinctly across reviewers, rather than uniformly increasing them.

However, substantial heterogeneity remains when reviewers are compared at a common workload and vacancy reference. On the same 122,536 applications to 2,524 postings, adding workload and its interaction with perceived origin reduces the sampling-error-adjusted standard deviation only from 1.81 to 1.75 percentage points, and we still reject equality of reviewers' gaps ($p < 0.001$; Appendix Figure G5). .

Reviewer heterogeneity also remains substantial after controlling for observed prior experience ($p < 0.001$; Appendix Figures G6 and G7). These measure observed assignment histories rather than employment tenure.

Application order. Finally, we investigate whether the ethnic gap varies with an application's arrival position within a vacancy. Its interaction with perceived origin is not statistically significant in either linear or quadratic specifications (Appendix Tables F7 and F8).

Figure F4: Observed heterogeneity in the shortlisting gap



Notes. This figure show regression estimates of equation 1 in which we separately interact the name indicator with occupation, customer contact, geographic area, or contract type. The dependent variable is the indicator for being shortlisted. Horizontal bars show 95 % confidence intervals with clustered standard errors at the posting level. Customer-facing roles are defined at the occupation level and include waiting and dining-room service, bar and counter service, reception, cashier roles, restaurant, bar and store management, and lifeguards.

Table F6: PPML robustness: heterogeneity across hiring contexts

Dimension	Separate specifications	Joint specification
Occupation	0.197	0.692
Customer contact	0.022	0.232
Geographic area	0.047	0.038
Contract	0.029	0.017

Notes. This table reports p-values from Wald tests based on PPML versions of Equation (1). The dependent variable is the indicator for being shortlisted. Each row tests whether the relative ethnic shortlisting gap is equal across categories of the indicated vacancy characteristic. Column (1) reports separate specifications that interact the immigrant-sounding-name indicator with one vacancy dimension at a time. Column (2) includes all four interaction blocks jointly.

Table F7: Application-order effects

	Baseline	Preferred	Quadratic
Immigrant-sounding name	-0.061^{***}	-0.037^{***}	-0.037^{***}
	(0.002)	(0.002)	(0.003)
Application order	-0.003	-0.016^{***}	-0.016^{***}
	(0.006)	(0.006)	(0.006)
Immigrant-sounding name × order	0.001	0.008	0.008
	(0.006)	(0.006)	(0.006)
Order^2			-0.001
			(0.019)
Immigrant-sounding name × order^2			0.002
			(0.023)
Job-posting and Month-Year FE	✓	✓	✓
Applicant and CV controls		✓	✓
Candidate–job match controls		✓	✓
WhatsApp assessment scores		✓	✓
Applications	123,311	123,311	123,311
Job postings	2,586	2,586	2,586
Mean Y (Italian-sounding name)	0.137	0.137	0.137
Adj. R^2	0.068	0.158	0.158

Notes. This table reports estimates of Equation (1), extended to examine application-order effects. The dependent variable is an indicator for being shortlisted. Application order measures arrival position within each posting, normalized from zero to one and centered at 0.5. Applications with identical timestamps receive their average rank. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table F8: Application-order contrasts

	Baseline	Preferred	Quadratic
Name gap: early application (0.10)	−0.062***	−0.040***	−0.041***
	(0.003)	(0.003)	(0.007)
Name gap: middle application (0.50)	−0.061***	−0.037***	−0.037***
	(0.002)	(0.002)	(0.003)
Name gap: late application (0.90)	−0.061***	−0.034***	−0.034***
	(0.003)	(0.003)	(0.004)
Late minus early: Italian-sounding name	−0.002	−0.013***	−0.013***
	(0.005)	(0.005)	(0.005)
Late minus early: immigrant-sounding name	−0.001	−0.007	−0.007
	(0.004)	(0.004)	(0.004)
Job-posting and Month-Year FE	✓	✓	✓
Applicant and CV controls		✓	✓
Candidate–job match controls		✓	✓
WhatsApp assessment scores		✓	✓
Applications	123,311	123,311	123,311
Job postings	2,586	2,586	2,586
Mean Y (Italian-sounding name)	0.137	0.137	0.137
Adj. R^2	0.068	0.158	0.158

Notes. This table reports estimates of Equation (1), extended to examine application-order effects. The dependent variable is an indicator for being shortlisted. Application order measures arrival position within each posting, normalized from zero to one and centered at 0.5. The first three rows report the immigrant-sounding-name gap at early (0.10), middle (0.50), and late (0.90) arrival positions. The last two rows compare shortlisting probabilities at late and early positions within each name group. $^{*}p < 0.10$, $^{**}p < 0.05$

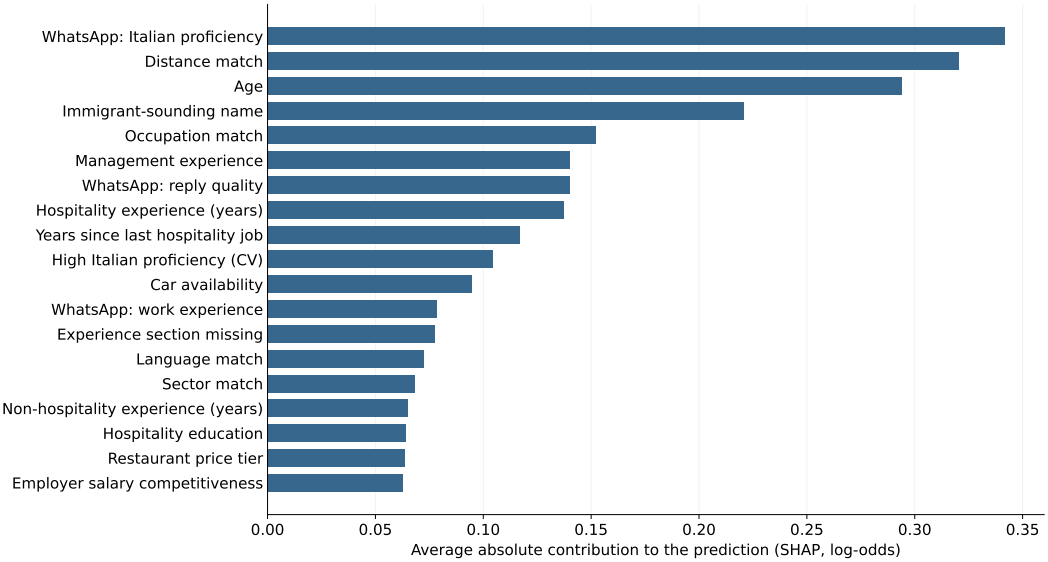
Table F9: Reviewer workload and the name gap

	(1)	(2)
Immigrant-sounding name	-0.0374^{***}	*Absorbed*
	(0.0023)	
Workload	0.0016	0.0011
	(0.0028)	(0.0030)
Immigrant-sounding name × Workload	-0.0046^{**}	-0.0035
	(0.0018)	(0.0025)
Job-posting and Month-Year FE	✓	✓
Applicant and CV controls	✓	✓
Candidate–job match controls	✓	✓
WhatsApp assessment scores	✓	✓
Application order control	✓	✓
Reviewer FE	✓	✓
Reviewer-specific name penalties		✓
Applications	123,311	123,311
Job postings	2,586	2,586
Mean Y (Italian-sounding name)	0.137	0.137
Adj. R^2	0.158	0.159

Notes. This table reports estimates of Equation (1), extended to examine workload effects. Workload counts applications arriving during the preceding seven days at the reviewer's other postings, excluding the current posting (leave-out). The dependent variable is an indicator for being shortlisted. Column (2) additionally allows the baseline name gap to differ across reviewers. $^{*}p < 0.10$, $^{**}p < 0.05$.

F.2 Predictive importance

Figure F5: Predictors of shortlisting under human screening



Notes. The vertical axis lists applicant and vacancy characteristics. The horizontal axis reports their predictive importance, measured by mean absolute SHAP values in log odds. We fit an XGBoost model to predict human shortlisting decisions and calculate SHAP values on applications from postings not used to train the model.

F.3 Flexible adjustment

Table F10: Human shortlisting with flexible adjustment

	Gap (pp)	Standard error
Preferred OLS	-3.743	0.227
Broad-controls OLS with posting FE	-3.851	0.230
Pooled cross-fitted DML	-3.828	0.261
Within-posting paired DML	-4.004	0.323

Notes. The outcome is an indicator for being shortlisted. Estimates and standard errors report the immigrant-sounding-name gap in percentage points. The preferred OLS uses the main specification. The second OLS specification uses posting fixed effects and the same predictors as the Double machine learning models. This predictor set is different from the one of the preferred specification. Double machine learning (DML) uses XGBoost first to predict both shortlisting and the name indicator, capturing nonlinearities and interactions. Then it regresses the shortlisting residual on the name-indicator residual.
We use five cross-fitting folds for prediction: applications to the same posting remain together, and each prediction comes from a model trained on other postings. Pooled DML does not include posting fixed effects. Within-posting paired DML differences the outcome and name indicator between randomly paired applications to the same posting, removing additive posting effects, and uses both applicants' characteristics to predict these differences. It uses 122,030 applications in 61,015 pairs, compared with 123,311 applications in the other specifications, and changes the weighting of observations. Standard errors are clustered by posting.

F.4 Post-hire comparisons in the Human-only sample

Table F11: Human-only post-hire outcomes: unadjusted name differences

	Survey separation			Worker reapplication			Employer reposting		
Days after hiring	30	60	90	30	60	90	30	60	90
Immigrant-sounding name	0.047	0.020	−0.009	0.002	0.026	0.059*	0.024	−0.019	0.027
	(0.086)	(0.091)	(0.098)	(0.021)	(0.027)	(0.032)	(0.034)	(0.036)	(0.044)
Controls	No	No	No	No	No	No	No	No	No
Hires	131	131	131	718	718	717	718	718	717
Mean Y (Italian-sounding name)	0.253	0.330	0.484	0.070	0.100	0.139	0.121	0.207	0.243
R^2	0.002	0.000	0.000	0.000	0.001	0.005	0.001	0.000	0.001

Notes. Outcomes are indicators for whether a hired worker's employment ended, the worker applied for another job on the platform, or the employer posted another vacancy in the same occupation within 30, 60, or 90 days of hiring. Separations are reported in employer surveys; worker reapplications and employer repostings are identified from platform records. Each coefficient reports the unadjusted difference between hires with immigrant-sounding and Italian-sounding names, in probability units. No controls are included. Standard errors clustered by employer are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table F12: Human-only post-hire outcomes: adjusted name differences

	Survey separation			Worker reapplication			Employer reposting		
Days after hiring	30	60	90	30	60	90	30	60	90
Immigrant-sounding name	0.063	0.030	0.008	0.018	0.049*	0.085***	0.030	−0.017	0.025
	(0.099)	(0.107)	(0.120)	(0.022)	(0.027)	(0.033)	(0.028)	(0.032)	(0.040)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hires	131	131	131	718	718	717	718	718	717
Mean Y (Italian-sounding name)	0.253	0.330	0.484	0.070	0.100	0.139	0.121	0.207	0.243
R^2	0.205	0.186	0.091	0.079	0.084	0.094	0.066	0.052	0.053

Notes. Outcomes are indicators for whether a hired worker's employment ended, the worker applied for another job on the platform, or the employer posted another vacancy in the same occupation within 30, 60, or 90 days of hiring. Separations are reported in employer surveys; worker reapplications and employer repostings are identified from platform records. Each coefficient reports the difference between hires with immigrant-sounding and Italian-sounding names, after controlling for the preferred controls. Standard errors clustered by employer are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

F.5 Missing information and the name gap

Table F13: Missing information and the shortlisting gap: separate interactions

Section	Gap: present	Gap: missing	Difference	SE	p	BH q
Experience	-3.68	-5.06	-1.37	0.47	0.003	0.017
Education	-3.86	-4.35	-0.49	0.43	0.250	0.416
Languages	-4.05	-3.95	0.10	0.41	0.806	0.806
Certifications	-3.37	-4.21	-0.84	0.49	0.090	0.224
Skills	-3.77	-4.16	-0.39	0.41	0.343	0.429

Notes. This table reports estimates of Equation (1). The dependent variable is an indicator for being shortlisted. Each row reports a separate specification adding an interaction between the immigrant-sounding-name indicator and an indicator for whether the corresponding profile section is missing. Column (1) report the estimated ethnic gap when the section is present, Column (2) when it is missing. Standard errors are clustered by posting. BH q-values apply the Benjamini–Hochberg correction across the five interaction tests.

Table F14: Missing information and the shortlisting gap: joint interactions

Section	Gap: present	Gap: missing	Difference	SE	p	BH q
Experience	-3.57	-5.44	-1.87	0.65	0.004	0.021
Education	-4.14	-3.70	0.43	0.60	0.473	0.591
Languages	-4.36	-3.55	0.81	0.48	0.091	0.228
Certifications	-3.54	-4.15	-0.61	0.52	0.241	0.401
Skills	-3.88	-4.09	-0.21	0.45	0.636	0.636

Notes. This table reports estimates of Equation (1). The dependent variable is an indicator for being shortlisted. All are estimated jointly. Column (1) report the estimated ethnic gap when the section is present, Column (2) when it is missing. Standard errors are clustered by posting. BH q-values apply the Benjamini–Hochberg correction across the five interaction tests.

Table F15: Does missing information moderate name gaps differently across reviewers?

Section	Reviewers	Applications	Postings	p	BH q
Experience	12	87,655	2,557	0.257	0.322
Education	12	87,655	2,557	0.185	0.308
Languages	12	87,655	2,557	0.009	0.044
Certifications	11	87,593	2,551	0.126	0.308
Skills	12	87,655	2,557	0.780	0.780

Notes. This table reports p-values from joint tests based on extensions of Equation (1). The dependent variable is an indicator for being shortlisted. Each row reports a separate specification interacting the immigrant-sounding-name indicator, an indicator for whether the corresponding profile section is missing, and reviewer identity, including all lower-order terms. Reviewers are included only if each name-group-by-missingness combination contains at least five applications, so reviewer counts and estimation samples vary across rows. Standard errors are clustered by posting. BH q-values apply the Benjamini–Hochberg correction across the five tests.

G Organizational heterogeneity: estimation and interpretation

G.1 Dispersion and empirical Bayes

Let $\widehat{\Delta}_r$ denote the estimated gap for reviewer r, and let s_r^2 be its estimated variance. The estimated sampling-noise-adjusted variance is therefore:

$$\widehat{\tau}^2 = \max\left\{ S^2_{\widehat{\Delta}} - \frac{1}{R}\sum_r s_r^2, 0 \right\}.$$

where $S^2_{\widehat{\Delta}}$ is the sample variance of the estimated reviewer gaps and R is the number of reviewers. Here, we set negative estimates to zero and we treat the covariances as negligible. When we use the full covariance matrix V, we instead subtract $\mathrm{tr}(PV)/(R-1)$, where $P = I - \mathbf{1}\mathbf{1}'/R$, to account for these covariances.

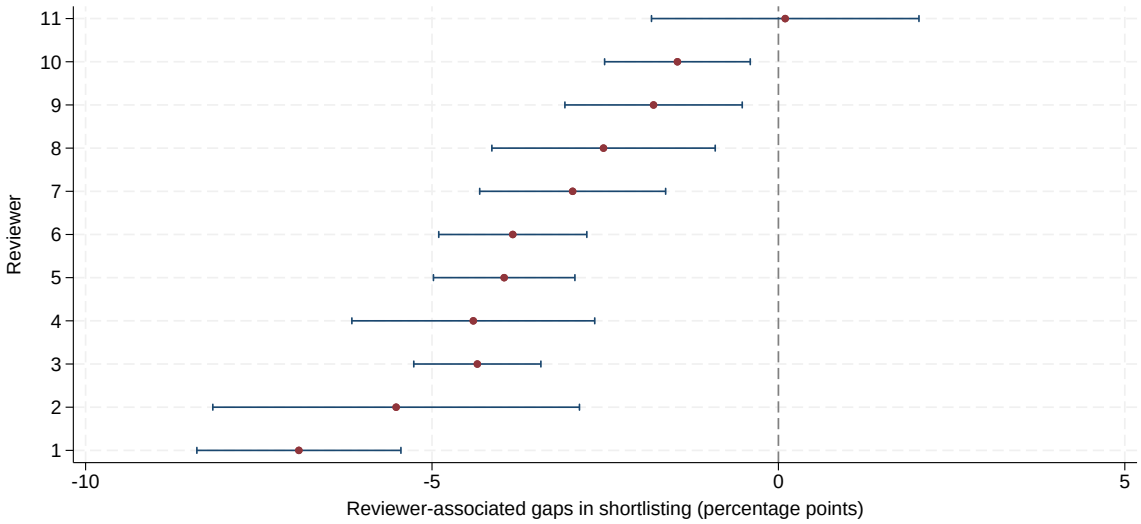
For the figures, we use empirical-Bayes estimates that pull each reviewer's estimated gap toward the equally weighted reviewer mean, $\bar{\Delta}$:

$$\widetilde{\Delta}_r = w_r\widehat{\Delta}_r + (1-w_r)\bar{\Delta}, \qquad w_r = \frac{\widehat{\tau}^2}{s_r^2 + \widehat{\tau}^2}.$$

Less precise estimates receive more shrinkage, while more precise estimates remain closer to their original values. Under the normal model, the posterior variance is $\widehat{\tau}^2 s_r^2/(\widehat{\tau}^2 + s_r^2)$. The plotted 95 percent posterior intervals use this variance, treating the estimated common

mean and variance as fixed.

Figure G1: Reviewer-associated shortlisting gaps: unshrunk estimates



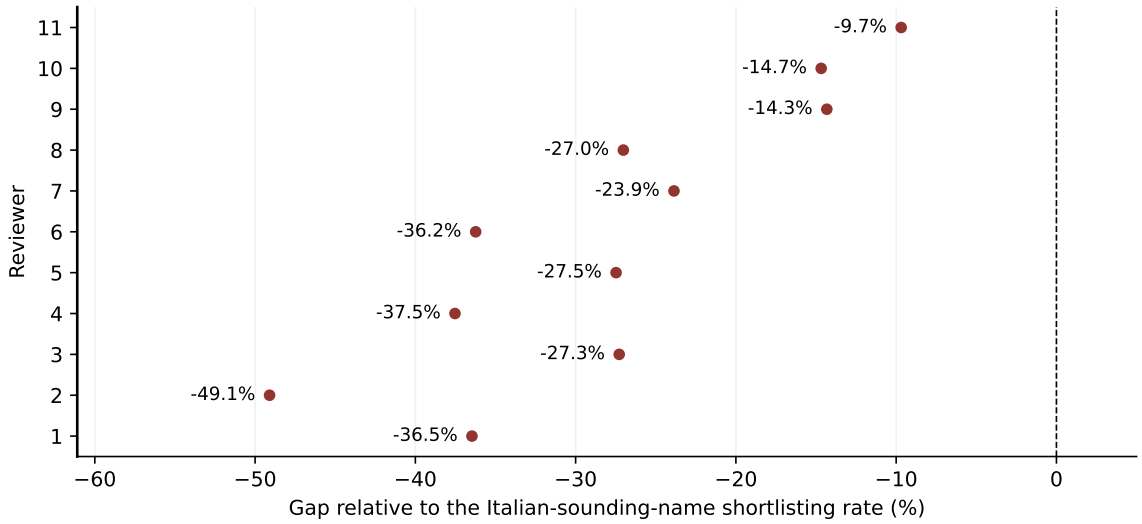
Notes. This figure reports reviewer-specific shortlisting gaps estimated using Equation (2). The vertical axis lists anonymized reviewers. The horizontal axis reports the estimated gap between applicants with immigrant-sounding and Italian-sounding names. No empirical-Bayes shrinkage applied. Horizontal bars show 95 percent confidence intervals with standard errors clustered at the posting level. Reviewer labels and ordering match Figure 4.

Table G1: Reviewer heterogeneity: PPML robustness

Reviewer	LPM full (pp)	LPM common (pp)	PPML gap (%)	95% CI
1	-6.92	-7.05	-39.02	[-45.01, -32.37]
2	-5.52	-5.97	-64.65	[-78.34, -42.31]
3	-4.35	-4.41	-30.04	[-35.74, -23.85]
4	-4.40	-4.46	-38.74	[-50.73, -23.84]
5	-3.96	-3.98	-35.55	[-42.13, -28.22]
6	-3.83	-3.93	-40.39	[-48.67, -30.78]
7	-2.97	-2.96	-25.17	[-35.75, -12.84]
8	-2.52	-2.66	-32.46	[-45.86, -15.74]
9	-1.80	-1.87	-12.71	[-22.27, -1.98]
10	-1.46	-1.61	-22.66	[-32.39, -11.53]
11	0.10	-0.37	-11.87	[-36.96, 23.19]

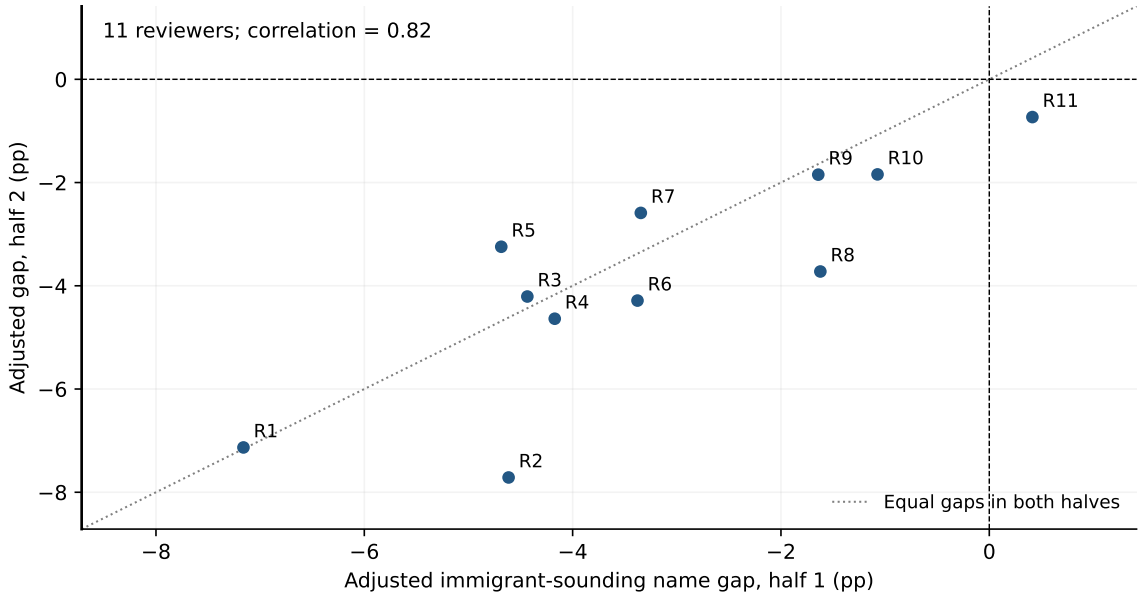
Notes. This table reports reviewer-specific shortlisting gaps using linear probability and PPML versions of Equation (2). The dependent variable is an indicator for being shortlisted. Each row reports estimates for one reviewer. Column (1) reports LPM estimates on the full sample (122,536 applications, 2,524 postings). Columns (2)–(4) use 118,583 applications and 2,311 postings. Column (2) reports LPM estimates on the same sample as PPML. Column (3) reports the PPML relative gap in percent, calculated as $100[\exp(\beta_r + \bar{Z}'\pi) - 1]$; Column (4) reports its 95 percent confidence interval. No empirical-Bayes shrinkage applied. Reviewer labels match Figure 4.

Figure G2: Reviewer gaps relative to the Italian-sounding-name rate



Notes. This figure reports reviewer-specific shortlisting gaps relative to the Italian-sounding-name shortlisting rate. Each estimated gap is divided by the corresponding reviewer's shortlisting rate for applicants with Italian-sounding names and multiplied by 100.

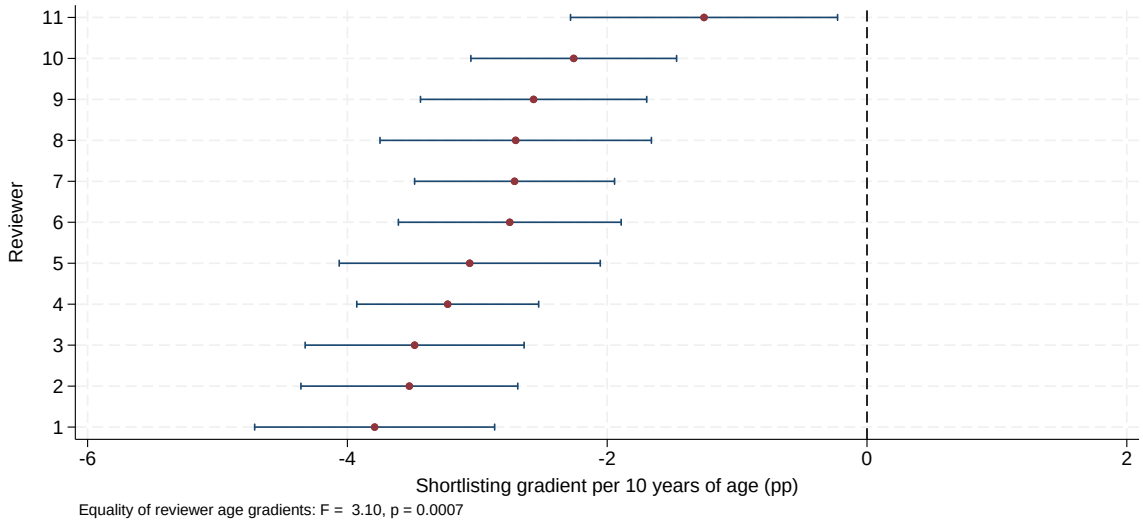
Figure G3: Reviewer shortlisting gaps in separate vacancy samples



Notes. This figure reports reviewer-specific ethnic shortlisting gaps estimated in two separate vacancy samples. Postings are split into two groups within each reviewer. The horizontal axis reports the adjusted gap in the first group; the vertical axis reports the adjusted gap in the second group. Each point represents one reviewer. The correlation between the estimates is 0.82.

G.2 Reviewer specific weights on applicant characteristics

Figure G4: Reviewer-specific age gradients



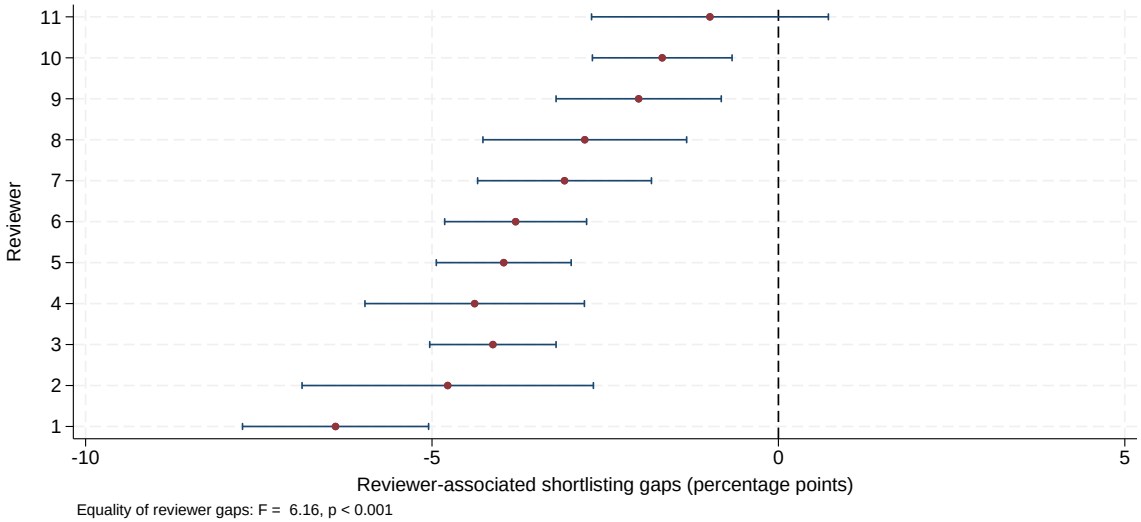
Notes. This figure reports reviewer-specific age gradients from Equation (2), controlling for an immigrant-sounding name indicator. The estimates measure the change in shortlisting probability associated with 10 additional years of age. Points report empirical-Bayes estimates. Bars show 95 percent posterior intervals, treating the estimated common mean and variance as fixed. The common sample contains 32,863 applications to 731 postings, covering eleven reviewers and 140 employers.

Table G2: Reviewer variation in the weights assigned to applicant characteristics

Characteristic	Reviewers	Mean	Raw SD	Adjusted SD	p	BH q
Immigrant sounding name	11	-3.38	1.66	1.22	0.000	0.000
Age	11	-3.32	1.03	0.56	0.001	0.005
Female	11	0.63	1.56	0.95	0.003	0.013
Hospitality experience	11	-0.14	1.70	0.29	0.000	0.000
Other experience	11	-0.14	0.84	0.16	0.010	0.034
Required occupation experience	11	-0.33	1.21	0.27	0.022	0.056
Employment gap	11	0.65	0.48	0.00	0.890	0.989
Years of education	11	0.15	0.14	0.00	0.995	0.995
Hospitality education	11	1.76	1.23	0.32	0.348	0.435
Foreign education	11	-1.32	1.03	0.33	0.328	0.435
Any certification	11	0.25	0.86	0.51	0.125	0.228
Occupation match	11	3.32	1.09	0.43	0.226	0.348
Experience match	11	3.29	1.53	0.69	0.219	0.348
Distance match	11	11.61	4.05	4.04	0.000	0.000
Sector match	11	1.99	0.33	0.00	0.995	0.995
Language match	11	1.39	1.91	1.24	0.012	0.035
Italian B2+	11	3.22	1.19	0.68	0.059	0.119
English B2+	11	0.82	0.62	0.00	0.859	0.989
CV structure	11	0.70	1.17	0.70	0.048	0.106
Profile picture	11	-0.07	0.79	0.23	0.341	0.435

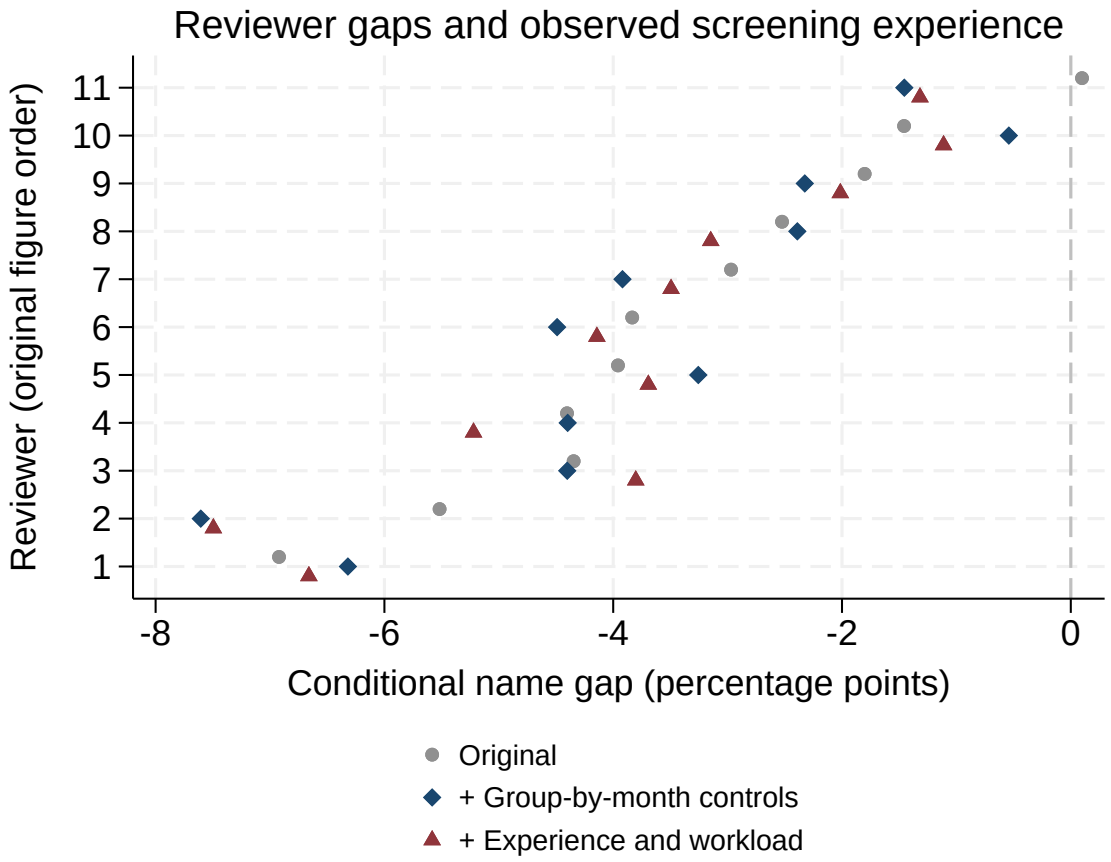
Notes. This table reports variation across eleven reviewers in the coefficients associated with applicant characteristics. We estimate a separate regression for each reviewer using the full preferred control set from Table 1, column (9). The dependent variable is an indicator for being shortlisted. Continuous-variable coefficients are scaled to a one-standard-deviation increase, using a common standard deviation calculated from observed values across reviewers. Binary-variable coefficients measure the difference between values of zero and one. Reported means and dispersion measures are expressed in percentage points. Standard errors in each regression are clustered by posting.

Figure G5: Reviewer heterogeneity before and after workload adjustment



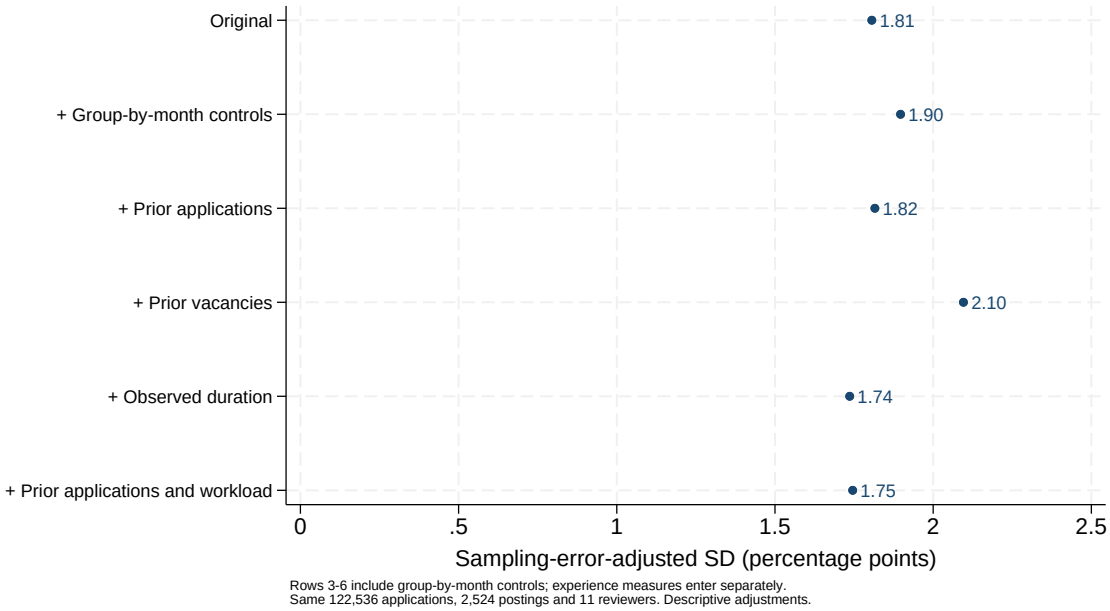
Notes. This figure compares reviewer-specific ethnic shortlisting gaps from Equation (2) before and after adding workload as a control and its interaction with the immigrant-sounding-name indicator. Workload measures applications arriving at the reviewer's other postings during the preceding seven days, transformed as log of one plus arrivals and standardized. We evaluate gaps using a common vacancy distribution and, in the adjusted model, a common workload reference. Both models use 122,536 applications to 2,524 postings and eleven reviewers. The sampling-noise-adjusted standard deviation is 1.81 percentage points before workload adjustment and 1.75 afterward. Points report empirical-Bayes estimates; bars show approximate 95 percent posterior intervals, treating the estimated common mean and variance as fixed. Tests of equality across reviewers use the unshrunk estimates.

Figure G6: Reviewer gaps before and after adjustment for observed experience



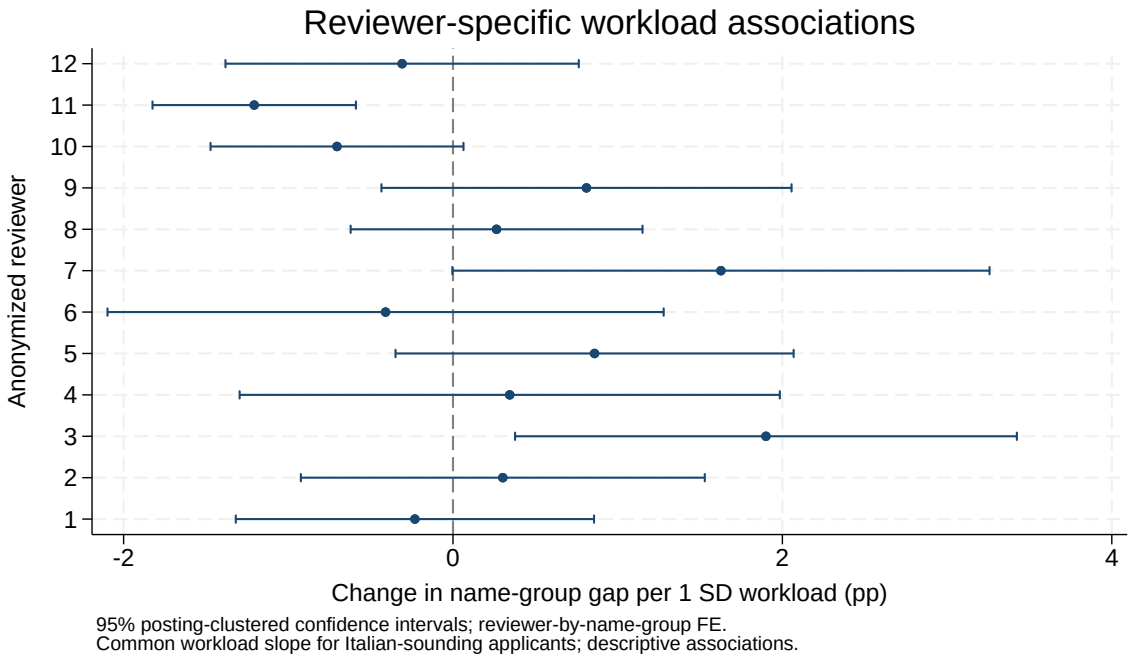
Notes. This figure compares reviewer-specific ethnic shortlisting gaps from Equation (2) before and after adjustment for reviewer experience and months. The calendar model adds interactions between the immigrant-sounding-name indicator and month. The red dots show estimated coefficients for equation with also standardized log prior application volume and workload, and their interaction with the name indicator. Prior application volume counts applications at the reviewer's other postings before the application day. Workload measures arrivals at other postings during the preceding seven days. The sample contains 122,536 applications, 2,524 postings, and eleven reviewers. We evaluate gaps at common covariate values. No empirical-Bayes shrinkage. Reviewer labels match Figure 4. The name-by-experience interaction is -0.50 percentage points per standard deviation (post-clustering standard error 0.63, $p = 0.425$).

Figure G7: Reviewer-gap dispersion under alternative experience adjustments



Notes. This figure reports reviewer-specific ethnic shortlisting gaps under alternative adjustments for reviewer experience with the sampling-noise-adjusted standard deviation. All specifications use the sample and baseline controls in Figure G6. Rows (3)–(6) add the indicated measures and their interactions with the immigrant-sounding-name indicator to the calendar model. We measure experience using prior applications, distinct prior postings, or days since the reviewer first appears in the records, each transformed as $\log(1+x)$ and standardized. Equality of reviewer-specific name slopes is rejected in every specification ($p < 0.001$).

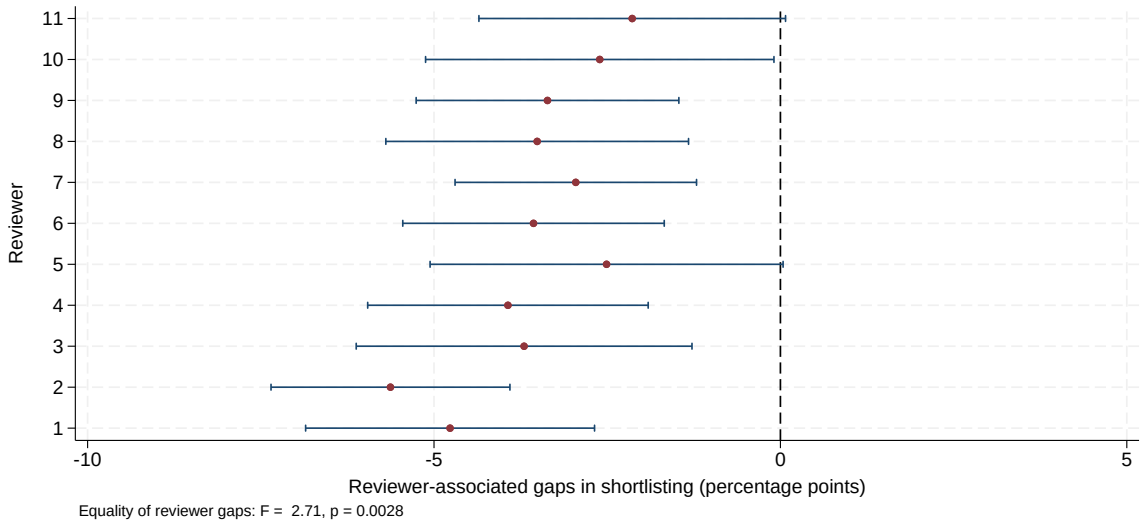
Figure G8: Reviewer-specific workload gradients in the ethnic gap



Coefficients measure changes in the immigrant-sounding minus Italian-sounding shortlisting gap per one SD of workload. Preferred controls and reviewer-specific baseline name gaps are included. The displayed specification uses a common workload slope for Italian-sounding applicants. Equality across reviewers has p=0.0004; allowing reviewer-specific Italian-sounding workload slopes gives p=0.0155. Reviewer 2 has only 227 applications across six postings.

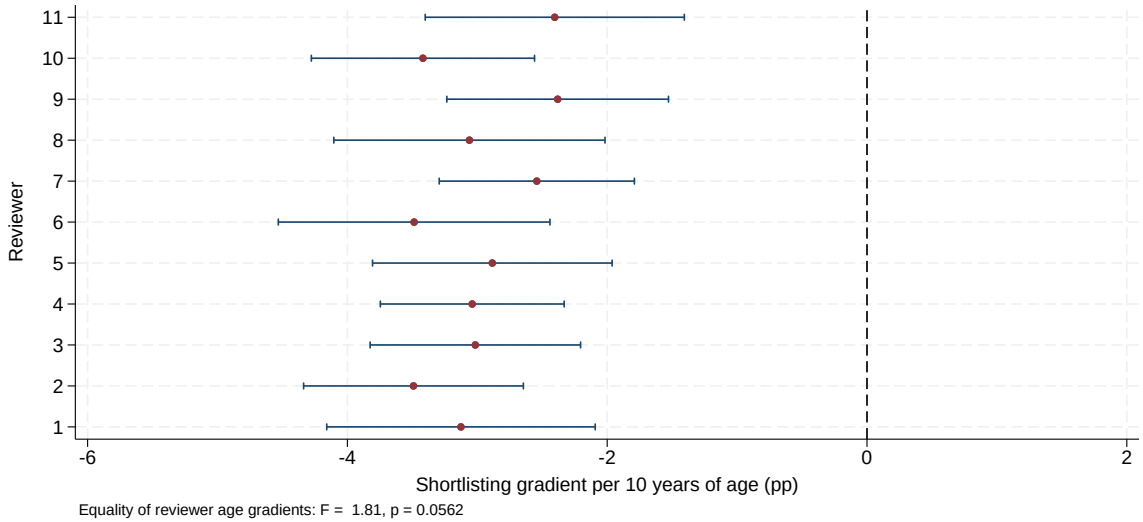
G.3 Joint reviewer and employer slopes

Figure G9: Reviewer gaps accounting for employer-specific gaps



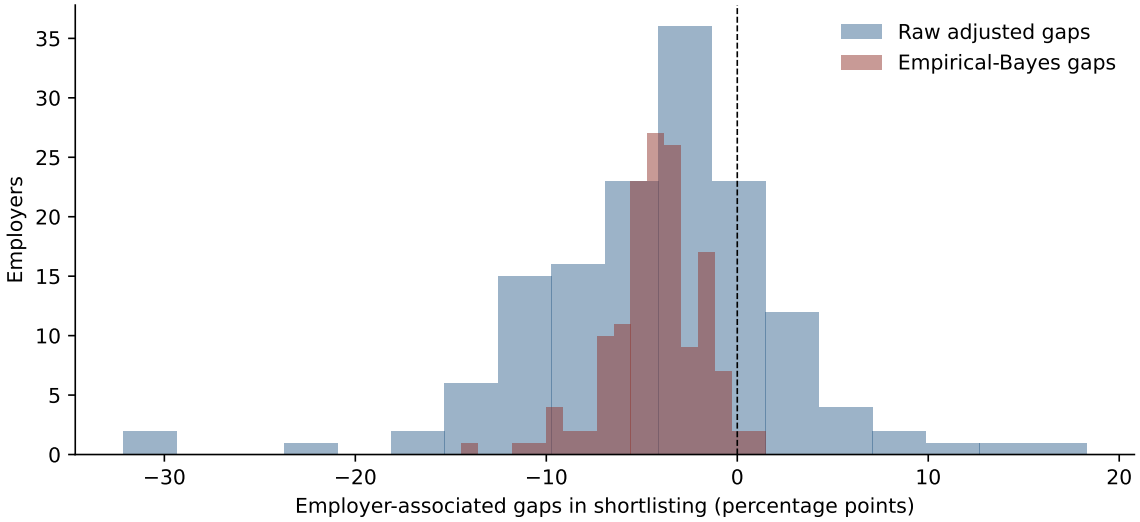
Notes. This figure reports reviewer-specific ethnic shortlisting gaps from Equation (3). The vertical axis lists anonymized reviewers; the horizontal axis reports gaps between applicants with immigrant-sounding and Italian-sounding names. Points report empirical-Bayes estimates; bars show 95 percent posterior intervals, treating the estimated common mean and variance as fixed. The sample contains 36,710 applications, 756 postings, eleven reviewers and 145 employers.

Figure G10: Reviewer age gradients accounting for employer-specific gradients



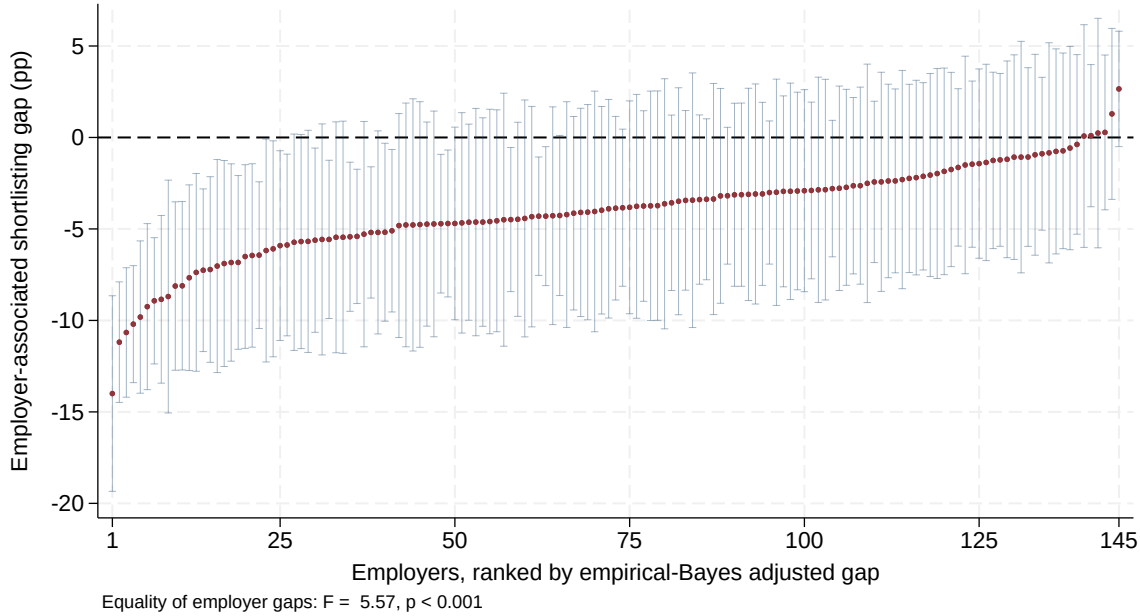
Notes. This figure reports reviewer-specific age gradients from a version of Equation (3) controlling for perceived origin. The vertical axis lists anonymized reviewers; the horizontal axis reports the change in shortlisting probability for ten additional years of age. Points report empirical-Bayes estimates; bars show 95 percent posterior intervals, treating the estimated common mean and variance as fixed. The sample contains 32,863 applications, 731 postings, eleven reviewers and 140 employers.

Figure G11: Employer-associated ethnic shortlisting gaps



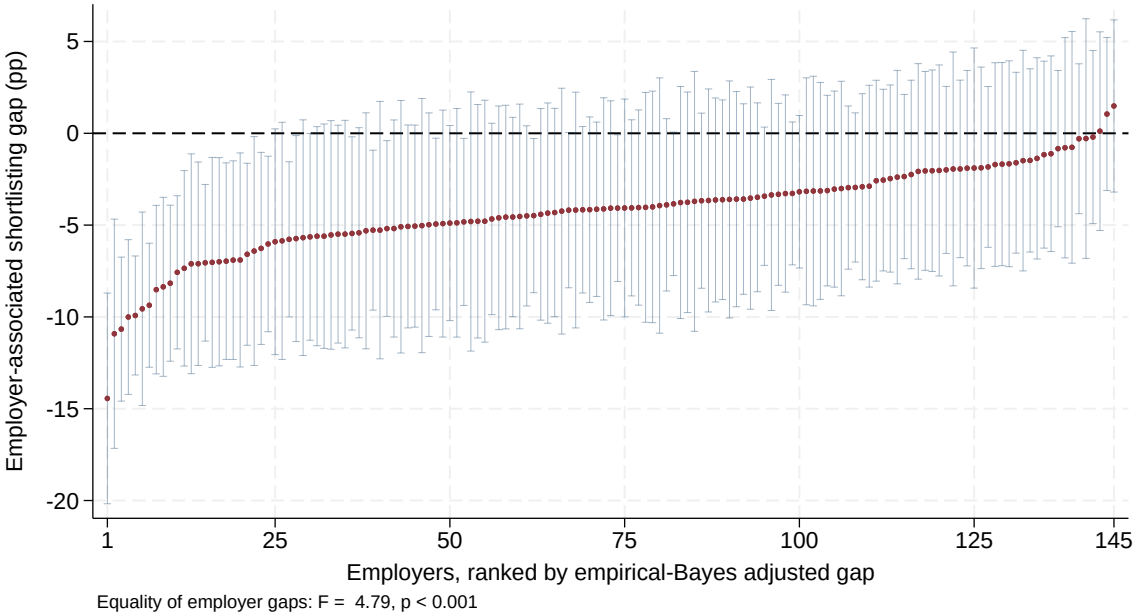
Employer-associated conditional gaps. These organizational associations need not reflect employer tastes; information, requirements, and reviewer adaptation can also contribute.

Figure G12: Employer gaps without reviewer-specific name slopes



Notes. This figure reports employer-specific ethnic shortlisting gaps from a version of Equation (3) that omits reviewer-specific name slopes. The horizontal axis ranks employers by their empirical-Bayes gap; the vertical axis reports gaps between applicants with immigrant-sounding and Italian-sounding names. Points report empirical-Bayes estimates; bars show approximate 95 percent posterior intervals, treating the estimated common mean and variance as fixed. The sample is the same as in Figure G13.

Figure G13: Employer gaps accounting for reviewer-specific gaps



Notes. This figure reports employer-specific ethnic shortlisting gaps from Equation (3). The horizontal axis ranks employers by their empirical-Bayes gap. The vertical axis reports gaps between applicants with immigrant-sounding and Italian-sounding names. Points report empirical-Bayes estimates; bars show 95 percent posterior intervals, treating the estimated common mean and variance as fixed. The sample contains 36,710 applications, 756 postings, eleven reviewers and 145 employers.

Figure G14: Employer age gradients accounting for reviewer-specific gradients



Notes. This figure reports employer-specific age gradients from Equation (3) that jointly estimates reviewer and employer age slopes, controlling for perceived origin. The horizontal axis ranks employers by their empirical-Bayes age gradient. The horizontal axis reports the change in shortlisting probability for ten additional years of age. Gradients are evaluated at a common reviewer and vacancy reference. Negative values indicate lower shortlisting probabilities for older applicants. Points report empirical-Bayes estimates; bars show 95 percent posterior intervals, treating the estimated common mean and variance as fixed. The sample contains 32,863 applications, 731 postings, eleven reviewers and 140 employers.

Table G3: Reviewer and employer variation in the joint model

	Immigrant-sounding name	Above-median age
Reviewer-associated SD (pp)	1.02	1.66
Employer-associated SD (pp)	4.37	5.60
Applications	32,863	32,863
Job postings	731	731
Supported reviewers	11	11
Supported employers	140	140

Notes. This table reports standard deviations of reviewer and employer shortlisting gaps from Equation (3). The dependent variable is an indicator for being shortlisted. Column (1) compares applicants with immigrant-sounding and Italian-sounding names. Column (2) compares applicants above and at or below the common median age of 31.54 years. The first row reports dispersion across reviewers after accounting for employer-specific gaps; the second reports dispersion across employers after accounting for reviewer-specific gaps. Standard deviations are adjusted for sampling noise using the full posting-clustered covariance matrix.

Table G4: Employer-associated gaps in the common joint-model sample

	Immigrant-sounding name	Above-median age
Applications	32,863	32,863
Job postings	731	731
Supported employers	140	140
Equally weighted mean gap (pp)	-4.42	-3.93
Raw cross-employer SD (pp)	7.48	7.80
Sampling-error-adjusted SD (pp)	4.37	5.60
p-value: equal employer slopes	< 0.001	< 0.001
Top-quintile share of estimated losses	56.9%	61.6%
Gini of estimated losses	0.541	0.606
Adj. R^2	0.152	0.147

Notes. This table summarizes employer-associated shortlisting gaps from Equation (3) and accounts for reviewer-specific gaps. The dependent variable is an indicator for being shortlisted. Column (1) compares applicants with immigrant-sounding and Italian-sounding names; column (2) compares applicants above and at or below the common median age of 31.54 years. Both columns use the same sample. Standard deviations are adjusted for sampling noise using the full posting-clustered covariance matrix. The p-values test equality of employer-specific group slopes. Estimated losses multiply each employer's negative empirical-Bayes gap by its number of applications in the corresponding group. The top-quintile share reports the proportion of these losses concentrated among the 20 percent of employers with the largest estimated losses. The Gini measures how concentrated these losses are across employers.

G.4 Crossed pair-gap decomposition

We first re-estimate Equation (3), replacing the separate reviewer and employer group slopes with a slope specific to each reviewer–employer pair:

$$Y_i = \alpha_j + \gamma_{r(i),t} + X_i'\lambda + M_i'\theta + G_i\left[d_{r(i),e(j)} + Z_j'\pi\right] + \varepsilon_i. \tag{6}$$

The dependent variable is an indicator for being shortlisted. Controls and fixed effects are retained from the joint model. For the age comparison, we replace the name indicator with age measured in ten-year units and control for perceived origin; the ethnic specification controls for age. We evaluate each pair's gap or gradient at the same vacancy reference, obtaining $\widehat{\Delta}_{re} = \widehat{d}_{re} + \bar{Z}'\widehat{\pi}$.

In the second step, we model Δ_{re} as:

$$\widehat{\Delta}_{re} = \mu + a_r + b_e + q_{re} + u_{re}. \tag{7}$$

Where a_r is the reviewer component shared across employers served by reviewer r, and b_e is the employer component shared across reviewers serving employer e. The component q_{re}

is the remaining variation specific to their pairing, while u_{re} is the estimation error in the first-stage gap.

We estimate the component variances under a crossed normal random-effects model:

$$\mathrm{Var}(\widehat{\boldsymbol{\Delta}}) = \sigma_r^2 Z_r Z_r' + \sigma_e^2 Z_e Z_e' + \sigma_q^2 I + V,$$

where Z_r and Z_e link pairs to reviewers and employers, and V is the estimated posting-clustered covariance matrix of the first-stage gaps. The reviewer, employer, and pair components are assumed to be mutually independent, mean-zero normal random effects, independent of first-stage estimation error. We treat V as fixed when estimating the variance components.

Table G5: Crossed decomposition on common demographic samples

	Reviewer	Employer	Pair
Continuous-age common sample; sampling covariance included			
Immigrant-sounding name	1.69	1.21	4.43
Age per ten years	0.44	2.17	1.39

Notes. This table reports estimated standard deviations of the random components in Equation (7). Columns (1), (2), and (3) report the reviewer, employer, and pair components, respectively. The first row uses conditional ethnic shortlisting gaps; the second uses shortlisting gradients per ten additional years of age. Both rows use the same 16,812 applications, 361 postings, 42 employers, and 87 reviewer–employer pairs. Each pair has at least two postings and enters the decomposition once. Estimates account for the full posting-clustered sampling covariance matrix of the first-stage gaps, treated as fixed.

G.5 Leave-posting-out variance correction

We assess whether the estimated dispersion changes when we correct for sampling noise using the leave-posting-out approach of Kline, Saggio and Sølvsten (2020). For the joint model in Equation (3), we write

$$\mathrm{Var}_w(\delta_r + \eta_e) = \mathrm{Var}_w(\delta_r) + \mathrm{Var}_w(\eta_e) + 2\,\mathrm{Cov}_w(\delta_r, \eta_e).$$

We give equal weight to each observed reviewer–employer pair. We leave out one posting at a time, re-estimate the model, and use the omitted observations to correct the variance and covariance estimates for sampling noise. The correction still requires errors to be independent across postings and the conditional mean to be correctly specified. We apply it to the unshrunk estimates.

Table G6: Leave-posting-out decomposition on supported employer samples

	Reviewer	Employer	Sorting	Total
Name: variance (pp^2)	1.28	7.15	-0.27	8.16
Name: SD (pp)	1.13	2.67	—	2.86
Age per ten years: variance (pp^2)	0.28	1.05	0.49	1.82
Age per ten years: SD (pp)	0.53	1.02	—	1.35

Notes. This table reports leave-posting-out estimates of the variance components from Equation (3). Columns (1) and (2) report reviewer and employer variances. Column (3) reports twice their weighted covariance. Column (4) reports the total. We keep employers with at least five postings and 30 applications fir each name group.

H Flexible same-applicant comparisons

H.1 Agreement at equal screening capacity

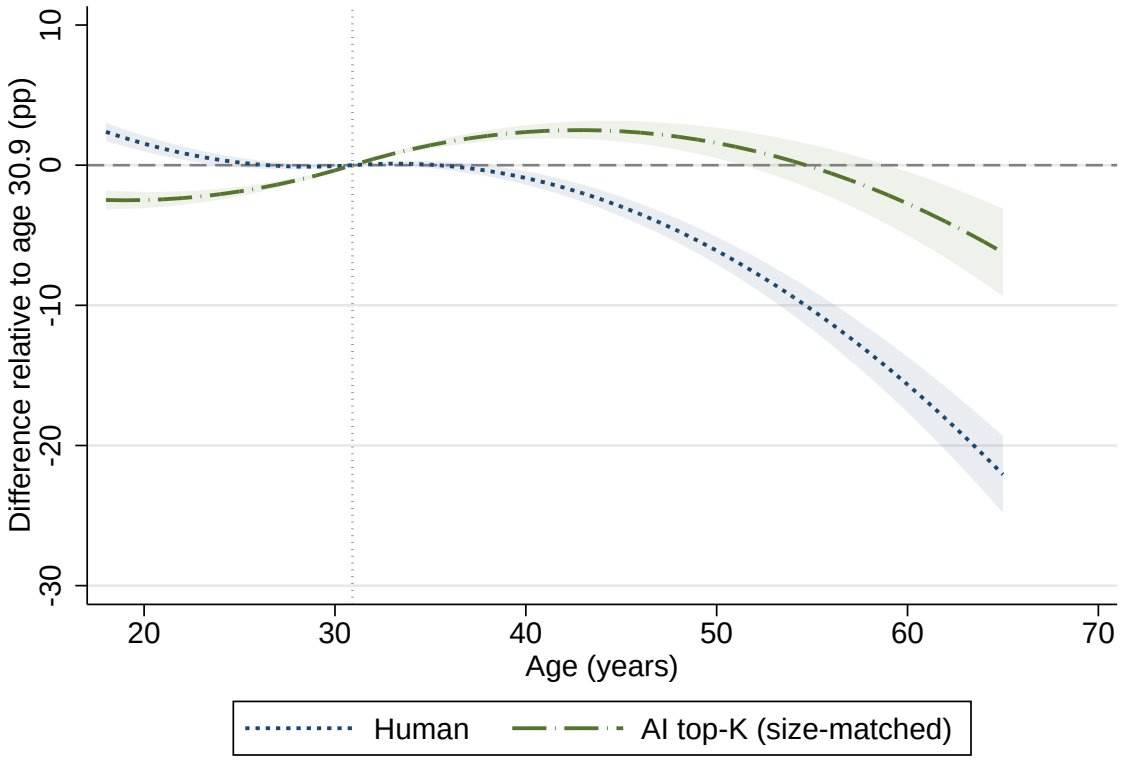
Table H1: Human and AI agreement at the same shortlist size

	Italian-sounding name		Immigrant-sounding name		All	
	N	%	N	%	N	%
Both shortlist	1012	10.1%	440	4.9%	1452	7.6%
Human only	947	9.4%	519	5.8%	1466	7.7%
AI only	898	9.0%	568	6.3%	1466	7.7%
Neither	7175	71.5%	7485	83.1%	14660	77.0%
Applications	10032	100%	9012	100%	19044	100%
Overall agreement	81.6%		87.9%		84.6%	
Agreement among shortlisted	35.4%		28.8%		33.1%	
Human shortlist retained by AI	51.7%		45.9%		49.8%	

Notes. This table reports agreement between human and AI shortlisting decisions for the same applications. The sample contains 19,044 applications to 1,063 postings. Column (1) reports counts and Column (2) reports percentages for applicants with Italian-sounding names, columns (3) and (4) for immigrant-sounding names, and columns (5) and (6) for all applicants. Overall agreement includes applications selected by both rules or neither rule. Agreement among shortlisted applicants is the share selected by both among those selected by at least one rule. We exclude applications with unclassified names and postings where humans select none or all applicants before constructing the AI shortlist.

H.2 Comparing conditional selection associations

Figure H1: Human and AI screening: quadratic age profiles



Notes. This figure reports conditional age profiles for human and AI shortlisting decisions on the same applicant pool, including linear and quadratic age terms and the joint applicant and match controls. The horizontal axis reports age in years; the vertical axis reports the estimated difference in shortlisting probability relative to age 30.9. 95 percent confidence intervals.

H.3 Flexible adjustment

Table H2: Human and AI: flexible adjustment for the name gap

	Human	AI	AI minus Human	Applications
OLS, pooled	-6.80^{***}	-4.33^{***}	2.47^{***}	19,044
	(0.56)	(0.55)	(0.57)	
OLS, posting FE	-6.39^{***}	-3.56^{***}	2.84^{***}	19,044
	(0.56)	(0.55)	(0.65)	
Pooled DML	-5.74^{***}	-2.86^{***}	2.88^{***}	19,044
	(0.57)	(0.55)	(0.61)	
Paired DML	-6.30^{***}	-2.62^{***}	3.68^{***}	18,528
	(0.77)	(0.74)	(0.91)	

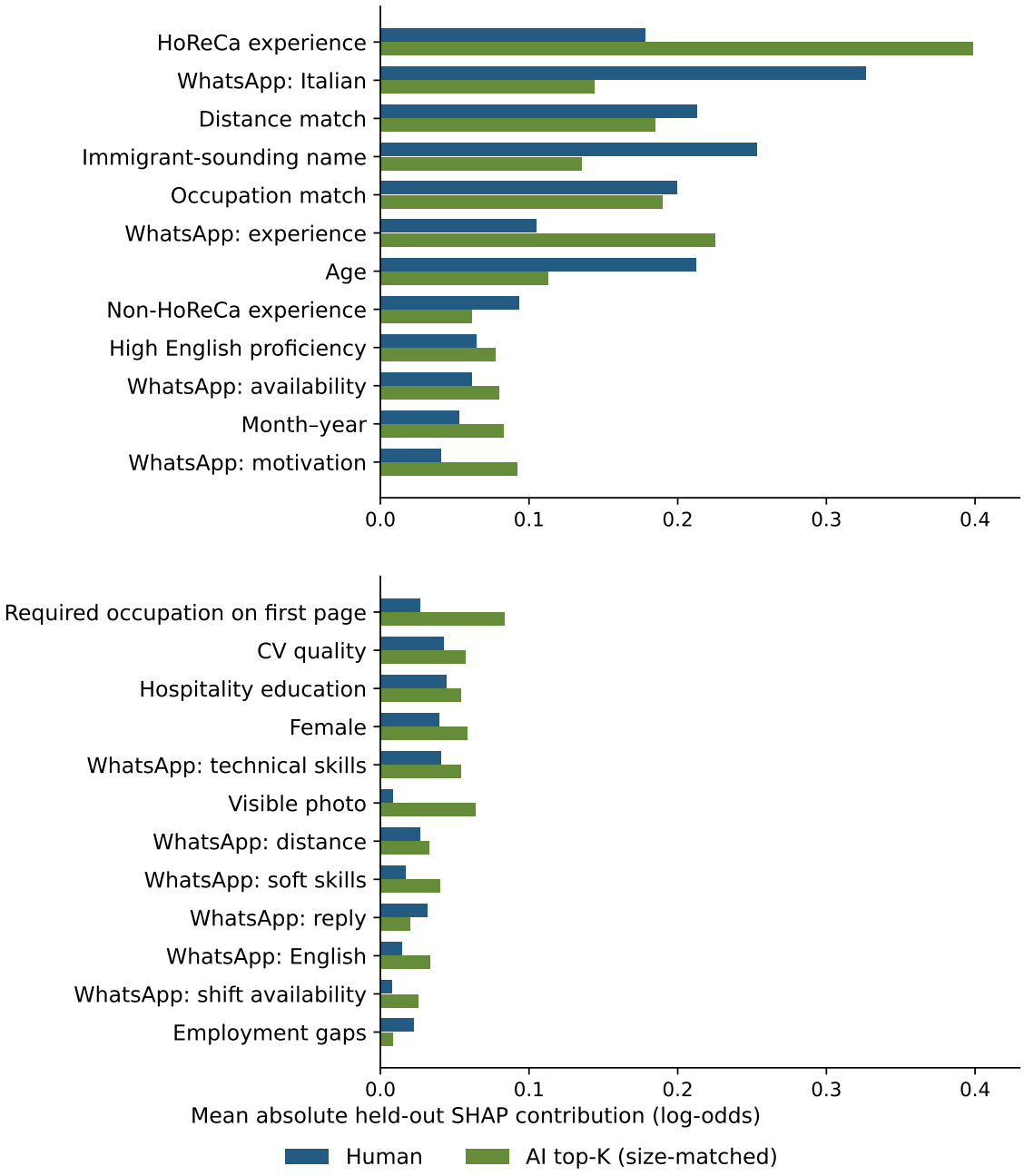
Notes. This table reports estimated ethnic shortlisting gaps under human and AI screening using alternative adjustment methods. The dependent variable is an indicator for being shortlisted for each corresponding rule. Columns (1) reports the gaps for human Column (2) reports the gap under AI screening, column (3) reports their difference, and column (4) reports the number of applications. The first two rows use OLS, without and with posting fixed effects. The last two rows use pooled and within-posting paired double machine learning (DML), allowing flexible adjustment for observed characteristics. All specifications use the common broad control set, including continuous age and calendar indicators, and cover 1,063 postings. Standard errors clustered at the posting level. $p < 0.10$, $p < 0.05$, $p < 0.01$

Table H3: Human and AI: flexible age-gap adjustment

	Human	AI top-K_j	AI minus Human	Applications	Postings
OLS pooled	-1.52^{**}	$+4.48^{***}$	$+5.99^{***}$	16,289	1,063
	(0.63)	(0.58)	(0.63)		
OLS posting FE	-2.11^{***}	$+4.83^{***}$	$+6.94^{***}$	16,283	1,057
	(0.63)	(0.59)	(0.71)		
Pooled DML	-2.15^{***}	$+0.52$	$+2.67^{***}$	16,289	1,063
	(0.73)	(0.66)	(0.76)		
Within-posting paired DML	-0.86	$+2.89^{***}$	$+3.75^{***}$	15,762	1,057
	(1.00)	(0.94)	(1.14)		

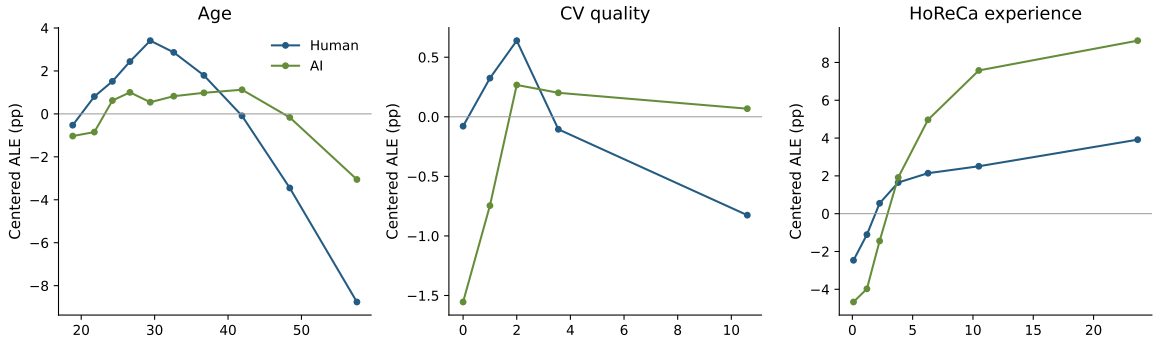
Notes. This table reports estimated shortlisting gaps under human and AI screening between applicants above and at or below the common median age of 30.9 years. The dependent variable is an indicator for being shortlisted for each corresponding rule. We use flexible adjustment for observed characteristics. The sample includes applications with no missing age. Age is excluded from the predictors, while the immigrant-sounding-name indicator is included. Standard errors clustered at the posting level. $p < 0.10$, $p < 0.05$, $p < 0.01$

Figure H2: Human and AI: predictive importance



Notes. This figure reports the predictive importance of applicant characteristics in separate models of human and AI shortlisting decisions. Importance is measured in log-odds with the mean absolute SHAP contribution.

Figure H3: Human and AI: nonlinear predictive profiles



Notes. This figure reports nonlinear predictive profiles from separate XGBoost models of human and AI shortlisting decisions. The horizontal axis in each panel shows the value of the indicated characteristic. The vertical axis shows its local effect on the model's prediction. Higher values indicate higher predicted shortlisting relative to the centered reference. The models include the immigrant-sounding-name indicator as a predictor.

H.4 Reviewer gaps on a common vacancy distribution

For each screening rule $s \in \{H, A\}$, we estimate

$$Y_i^s = \alpha_j^s + \gamma_{r(i),t}^s + X_i'\lambda_s + M_i'\theta_s + G_i\left[\beta_{r(i)}^s + Z_j'\pi_s\right] + \varepsilon_i^s. \tag{8}$$

i = application, j = job posting, $r(i)$ = reviewer, and t = month-year. The dependent variable is an indicator for being shortlisted. We use the same controls as Equation (4), with posting and reviewer-by-month fixed effects. We interact the immigrant-sounding-name indicator G_i with reviewer identity and vacancy characteristics Z_j. Standard errors are clustered at the posting level.

The sample contains postings with both name groups and occupations non-missing, and reviewers with at least five postings and 30 applications from each name group. The common sample contains 18,685 applications to 1,001 postings assigned to nine reviewers. Reported gaps are $\widehat{\Delta}_r^s = \widehat{\beta}_r^s + \overline{Z}'\widehat{\pi}_s$, where $\overline{Z}$ is the application-weighted vacancy distribution of this sample, used for every reviewer and both rules.

H.5 Workload under the two screening rules

Table H4: Workload and name gaps under Human and AI screening

Dependent variable	Human Shortlisted	AI Shortlisted	AI minus Human Difference in shortlisting
A. Common name gap			
Immigrant-sounding name	−5.733***	−3.581***	2.153***
	(0.545)	(0.532)	(0.635)
Workload	0.722	−0.868	−1.590**
	(0.627)	(0.595)	(0.705)
Immigrant-sounding name × workload	−0.689	−0.320	0.369
	(0.462)	(0.447)	(0.534)
B. Reviewer specific name gaps			
Workload	0.590	−0.955	−1.545**
	(0.648)	(0.618)	(0.737)
Immigrant-sounding name × workload	−0.407	−0.143	0.263
	(0.529)	(0.501)	(0.630)
Job-posting and month-year FE	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes
Applicant, CV, match and WhatsApp controls	Yes	Yes	Yes
Applications	19,044	19,044	19,044
Job postings	1,063	1,063	1,063

Notes. This table reports the association between reviewer workload and ethnic shortlisting gaps under human and AI screening. The dependent variable is an indicator shortlisting. Workload measures applications arriving at the reviewer's other postings during the preceding seven days, transformed as $\log(1 + \text{workload})$ and standardized. Panel B additionally includes reviewer-specific name gaps. Within each posting, AI selects the same number of applicants as human reviewers, allowing fractional selection probabilities at cutoff ties. Coefficients and standard errors are in percentage points per one standard deviation increase in log workload. Standard errors clustered by posting are reported in parentheses. Estimates describe associations at observed human workload rather than causal effects of changing workload or deploying AI. $p < 0.10$, $p < 0.05$, $p < 0.01$.

I AI-Hybrid versus Human + Information

Table I1: AI-Hybrid versus Human + Information: Balance

	Pooled				CEM			
	Human + Information	AI-Hybrid	Std. diff.	p	Human + Information	AI-Hybrid	Std. diff.	p
Panel A. Job-posting characteristics								
Any language required	0.319	0.277	-0.093	0.351	0.272	0.244	-0.060	0.635
Italian required	0.010	0.015	0.037	0.616	0.005	0.017	0.101	0.164
Required languages	0.346	0.292	-0.105	0.267	0.290	0.256	-0.066	0.560
Maximum distance	25.676	23.856	-0.118	0.377	23.650	22.982	-0.043	0.710
Local residence required	0.029	0.003	-0.211	0.192	0.005	0.005	0.000	1.000
Minimum experience	1.919	1.646	-0.241	0.013	1.579	1.602	0.021	0.848
Senior / management role	0.139	0.103	-0.112	0.211	0.092	0.100	0.022	0.810
Customer-facing role	0.481	0.476	-0.010	0.916	0.517	0.488	-0.057	0.682
Monthly net salary	1734.619	1647.372	-0.165	0.103	1636.482	1659.481	0.043	0.622
Permanent contract	0.096	0.138	0.133	0.267	0.055	0.124	0.216	0.018
Full-time	0.919	0.872	-0.151	0.116	0.924	0.896	-0.093	0.409
Days per week	5.680	5.516	-0.275	0.007	5.627	5.576	-0.086	0.541
Split shifts	0.137	0.170	0.091	0.347	0.106	0.180	0.205	0.093
Seasonal position	0.171	0.089	-0.246	0.018	0.142	0.071	-0.213	0.061
Number of benefits	2.156	2.287	0.130	0.231	2.341	2.230	-0.110	0.501
Panel B. Employer characteristics								
Hotel / resort	0.142	0.158	0.043	0.719	0.101	0.116	0.041	0.691
Formal restaurant	0.102	0.065	-0.131	0.226	0.138	0.101	-0.136	0.464
Staff size	11.494	12.503	0.106	0.360	13.121	11.482	-0.173	0.447
New employer	0.243	0.181	-0.152	0.156	0.210	0.175	-0.085	0.546
Panel C. Applicant-pool characteristics								
Mean age	35.371	35.664	0.073	0.422	35.042	35.238	0.049	0.750
Share female	0.280	0.293	0.075	0.465	0.269	0.286	0.099	0.300
Mean hospitality experience	6.049	5.956	-0.011	0.902	5.656	6.225	0.066	0.612
Mean experience in required occupation	0.443	0.489	0.079	0.344	0.422	0.464	0.073	0.535
Job postings	210	393			136	211		

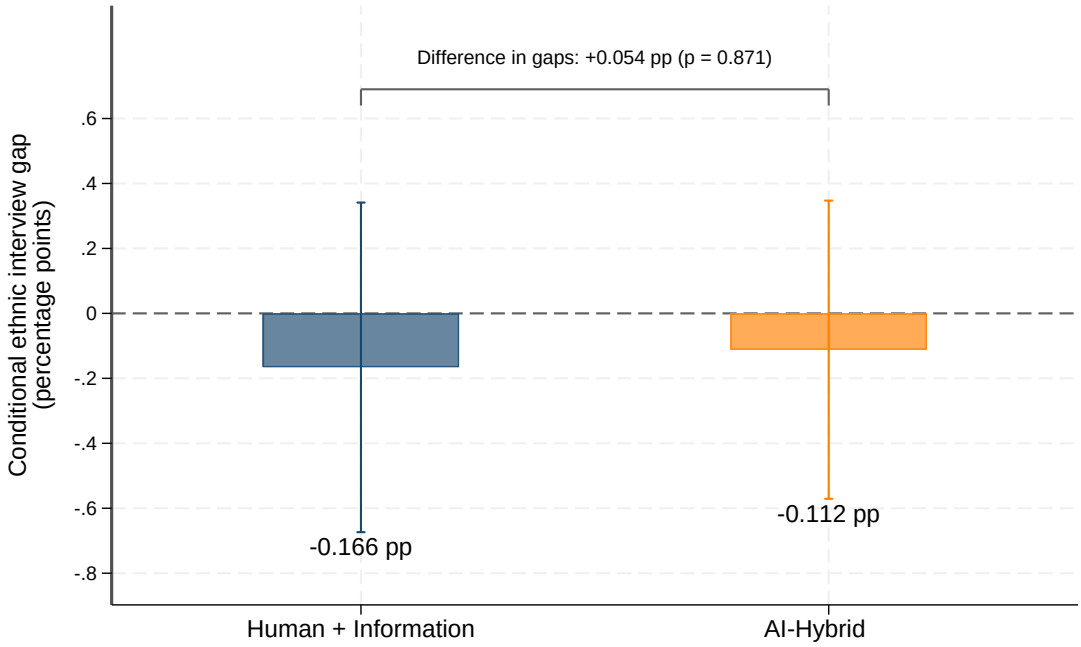
Notes. This table compares job-posting, employer, and applicant-pool characteristics between Human + Information and AI-Hybrid in July–September 2025. Columns (1) and (2) report unweighted posting-level means; columns (3) and (4) report the standardized difference and the p-value for equality of means. Columns (5)–(8) report the corresponding statistics for matched postings, using CEM weights. CEM p-values come from regressions with standard errors clustered by employer. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table I2: AI-Hybrid vs Human + Information

	Pooled		CEM	
	Interviewed	Hired	Interviewed	Hired
AI-Hybrid	0.00553	−0.00166	0.00609	−0.00290
	(0.00388)	(0.00173)	(0.00612)	(0.00273)
Immigrant-sounding name	−0.00166	−0.00328*	−0.00877*	−0.00707**
	(0.00258)	(0.00174)	(0.00471)	(0.00316)
AI-Hybrid × Immigrant-sounding name	0.00054	0.00380**	0.00325	0.00666**
	(0.00336)	(0.00190)	(0.00677)	(0.00326)
Month fixed effects	Yes	Yes	Yes	Yes
Applicant and CV controls	Yes	Yes	Yes	Yes
Candidate–job match controls	Yes	Yes	Yes	Yes
WhatsApp assessment scores	Yes	Yes	Yes	Yes
Applications	31,198	31,198	16,392	16,392
Job postings	603	603	353	353
R^2	0.0310	0.0112	0.0451	0.0151

Notes. This table reports regression estimates of Equation (5), using an indicator for being interviewed as the dependent variable in columns (1) and (3), and an indicator for being hired in columns (2) and (4). Columns (1) and (2) use all applications in the July–September 2025 comparison sample and give equal weight to each application. Columns (3) and (4) use the CEM matched sample, with each application receiving its posting's matching weight. Human + Information is the reference regime. Standard errors are clustered at employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure I1: AI-Hybrid versus Human + Information: the conditional interview gap



Notes. This figure reports estimates from Equation (5), using an indicator for being interviewed as the dependent variable. The horizontal axis shows the screening regime. The vertical axis reports the conditional interview probability gap between applicants with immigrant-sounding and Italian-sounding names. Bars show β under Human + Information and $\beta + \delta$ under AI-Hybrid. The bracket reports the change in the gap (δ) and its p-value. Estimates use 31,198 applications to 603 postings in July–September 2025, with equal weight given to each application, month fixed effects, and applicant, CV, match, and WhatsApp controls. 95 percent confidence intervals and standard errors clustered at the employer level.

Table I3: AI-Hybrid vs Human + Information: robustness of the ethnic hiring-gap estimate

Specification	Change in ethnic hiring gap (pp)	p-value
Baseline: applicant controls and month FE	0.380	0.046
Posting and month FE	0.383	0.068
Reviewer-by-month FE	0.356	0.059
Posting FE + name × new employer	0.384	0.067
Posting FE + name × four posting characteristics	0.404	0.060
Matched sample, unweighted	0.557	0.063
Matched sample, CEM weights	0.666	0.043
Matched sample, posting-normalized CEM weights	0.694	0.142
Mechanism sample, CEM weights	0.723	0.032

Notes. This table reports robustness checks for Equation (5). The dependent variable is an indicator for being hired. Each row reports a separate specification. The estimate column reports the coefficient on AI-Hybrid × immigrant-sounding name. All specifications retain the controls of (5). The four posting characteristics are new employer status, seasonality, minimum experience requirements, and days per week. The pooled sample contains 31,198 applications and 603 postings. The matched sample contains 16,392 applications and 353 postings. Under CEM weighting, each application receives its posting's matching weight. Posting-normalized CEM divides this weight by the posting's number of applications. Standard errors are clustered at employer level.

Table I4: AI-Hybrid vs Human + Information: age, interviews, and hiring

	Pooled		CEM	
	Interviewed	Hired	Interviewed	Hired
AI-Hybrid	0.00051	−0.00316	0.01337	0.00308
	(0.00816)	(0.00475)	(0.00860)	(0.00523)
Age (ten years)	−0.00445***	−0.00199*	−0.00424**	−0.00173
	(0.00161)	(0.00104)	(0.00175)	(0.00131)
AI-Hybrid × Age (ten years)	0.00103	0.00068	−0.00233	−0.00116
	(0.00184)	(0.00116)	(0.00204)	(0.00141)
Month fixed effects	Yes	Yes	Yes	Yes
Applicant and CV controls	Yes	Yes	Yes	Yes
Candidate–job match controls	Yes	Yes	Yes	Yes
WhatsApp assessment scores	Yes	Yes	Yes	Yes
Applications	26,047	26,047	13,677	13,677
Job postings	602	602	352	352
R^2	0.0291	0.0119	0.0409	0.0158

Notes. This table reports estimates of Equation (5) for age, controlling for perceived origin. The dependent variable is an indicator for being interviewed in columns (1) and (3), and for being hired in columns (2) and (4). Age is measured in ten-year units. Columns (1) and (2) use applications with observed age in the July–September 2025 comparison sample and give equal weight to each application. Columns (3) and (4) use the CEM matched sample, with each application receiving its posting's matching weight. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. .

Table I5: Components of employer satisfaction and willingness to reactivate the service

Outcome	Service	Assistance	Recommendation	Reactivate
AI-Hybrid	1.789***	1.332**	1.518***	0.210**
	(0.474)	(0.537)	(0.537)	(0.088)
Posting controls	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
Control mean	5.328	6.953	5.562	0.562
Observations	212	212	212	212
R^2	0.348	0.229	0.311	0.263

Notes. This table reports regression estimates using as dependent variables employers' satisfaction with the service in column (1), satisfaction with the assistance received in column (2), willingness to recommend the platform in column (3), and willingness to reactivate the service in column (4). The outcomes in columns (1)–(3) range from 0 to 10; reactivation is a binary outcome. Feedback is linked to the employer's nearest prior posting within 90 days. Standard errors clustered by employer are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table I6: Post-hire outcomes: AI-Hybrid versus Human + Information

Horizon	30 days	60 days	90 days	180 days
A. Survey-reported separation				
AI-Hybrid	0.061	0.212	0.121	0.030
	(0.338)	(0.339)	(0.336)	(0.332)
Control mean	0.273	0.455	0.545	0.636
Observed hires	35	35	35	35
B. Subsequent worker application				
AI-Hybrid	−0.039	−0.047	−0.054	−0.022
	(0.038)	(0.062)	(0.070)	(0.075)
Control mean	0.095	0.159	0.190	0.246
Observed hires	202	201	200	181
C. Employer reposting				
AI-Hybrid	0.149	0.165	0.089	0.099
	(0.106)	(0.118)	(0.124)	(0.128)
Control mean	0.111	0.175	0.270	0.328
Observed hires	202	202	202	185
D. Either administrative signal				
AI-Hybrid	0.102	0.094	0.029	0.105
	(0.107)	(0.118)	(0.124)	(0.127)
Control mean	0.206	0.317	0.413	0.475
Observed hires	202	202	202	187

Notes. This table reports regression estimates in which the dependent variable is a post-hire outcomes measured within 30 days in Column (1), 60 days in Column (2), 90 days in Column (3), and 180 days in Column (4). In Panel A, the dependent variable is an indicator for survey-reported separation. In Panel B, the dependent variable is a subsequent worker application. In Panel C, the dependent variable is employer reposting in the same occupation. In Panel D is either administrative. Panel A uses employers' survey responses; Panels B–D use platform records. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table I7: Post-hire outcomes by name group: AI-Hybrid vs Human + Information

Horizon	30 days	60 days	90 days	180 days
A. Survey-reported separation				
AI-Hybrid	0.416	0.411	0.411	0.432
	(0.294)	(0.300)	(0.300)	(0.292)
Immigrant-sounding name	0.500*	0.100	0.350	0.600**
	(0.279)	(0.370)	(0.343)	(0.244)
AI-Hybrid × Immigrant-sounding name	−0.747**	−0.532	−0.782*	−1.095***
	(0.309)	(0.425)	(0.402)	(0.330)
Control mean	0.000	0.400	0.400	0.400
Observations	31	31	31	31
B. Subsequent worker application				
AI-Hybrid	−0.023	0.017	0.002	0.024
	(0.047)	(0.065)	(0.076)	(0.084)
Immigrant-sounding name	0.046	0.152	0.107	0.240*
	(0.102)	(0.138)	(0.137)	(0.137)
AI-Hybrid × Immigrant-sounding name	−0.066	−0.200	−0.173	−0.270
	(0.112)	(0.150)	(0.151)	(0.163)
Control mean	0.095	0.119	0.167	0.195
Observations	181	180	179	165
C. Employer reposting				
AI-Hybrid	0.132	0.145	0.090	0.148
	(0.112)	(0.131)	(0.138)	(0.142)
Immigrant-sounding name	0.035	−0.058	−0.006	0.063
	(0.102)	(0.121)	(0.151)	(0.178)
AI-Hybrid × Immigrant-sounding name	0.001	0.047	−0.045	−0.191
	(0.129)	(0.149)	(0.178)	(0.205)
Control mean	0.095	0.190	0.286	0.317
Observations	181	181	181	169
D. Either administrative signal				
AI-Hybrid	0.115	0.128	0.079	0.154
	(0.113)	(0.129)	(0.134)	(0.131)
Immigrant-sounding name	0.082	0.094	0.146	0.224
	(0.136)	(0.183)	(0.201)	(0.173)
AI-Hybrid × Immigrant-sounding name	−0.110	−0.173	−0.255	−0.326
	(0.160)	(0.207)	(0.221)	(0.197)
Control mean	0.190	0.310	0.405	0.439
Observations	181	181	181	171

Notes. This table reports regression estimates in which the dependent variable is a post-hire outcomes measured within 30 days in Column (1), 60 days in Column (2), 90 days in Column (3), and 180 days in Column (4). In Panel A, the dependent variable is an indicator for survey-reported separation. In Panel B, the dependent variable is a subsequent worker application. In Panel C, the dependent variable is employer reposting in the same occupation. In Panel D is either administrative. Panel A uses employers' survey responses; Panels B–D use platform records. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

I.1 Mechanisms: Ranking, shortlist size, applicant characteristics, and employer decisions

Ranking and employer attention.

Table I8: AI-Hybrid versus Human + Information: shortlist ranking

	Pooled		CEM	
	Rank in shortlist (0–1)		Rank in shortlist (0–1)	
	(1)	(2)	(3)	(4)
AI-Hybrid	0.007	—	−0.015	—
	(0.016)		(0.026)	
Immigrant-sounding name	0.013	0.017	0.033	0.042
	(0.025)	(0.031)	(0.032)	(0.037)
AI-Hybrid × Immigrant-sounding name	−0.030	−0.023	−0.025	−0.041
	(0.031)	(0.039)	(0.044)	(0.053)
Control mean	0.498	0.498	0.489	0.489
Applicant controls	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
Posting controls	Yes	—	Yes	—
Posting fixed effects	No	Yes	No	Yes
CEM weights	No	No	Yes	Yes
Applications	2,995	2,995	1,567	1,567
Job postings	504	504	275	275

Notes. This table reports estimates of a version of Equation (5), using normalized shortlist rank as the dependent variable. Rank ranges from zero at the top to one at the bottom of the human shortlist or AI matched-sized shortlist. Columns (1) and (2) use the pooled comparison; columns (3) and (4) use the CEM matched comparison. Columns (1) and (3) include posting controls; columns (2) and (4) include posting fixed effects. All columns include month fixed effects and applicant, CV, preference, match, WhatsApp, timing, and application-order controls. Human ordering follows shortlisting timestamps. AI ordering follows descending scores. Standard errors clustered by employer are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table I9: AI-Hybrid versus Human + Information: employer outcomes controlling for score and ranking (pooled)

	Baseline	+ Score	+ Rank	+ Score and rank
A. Profile opened				
AI-Hybrid	-0.032	-0.034	-0.031	-0.032
	(0.036)	(0.037)	(0.036)	(0.036)
Immigrant sounding name	-0.042	-0.033	-0.043	-0.033
	(0.027)	(0.027)	(0.026)	(0.027)
AI-Hybrid × Immigrant sounding name	-0.004	-0.005	-0.002	-0.005
	(0.043)	(0.039)	(0.043)	(0.039)
R^2	0.138	0.159	0.139	0.160
B. Interviewed				
AI-Hybrid	-0.003	-0.004	-0.004	-0.004
	(0.021)	(0.021)	(0.021)	(0.021)
Immigrant sounding name	0.004	0.005	0.004	0.004
	(0.019)	(0.019)	(0.020)	(0.019)
AI-Hybrid × Immigrant sounding name	0.007	0.008	0.008	0.008
	(0.021)	(0.021)	(0.021)	(0.021)
R^2	0.044	0.048	0.045	0.049
C. Hired				
AI-Hybrid	-0.015	-0.015	-0.016	-0.016
	(0.010)	(0.010)	(0.010)	(0.010)
Immigrant sounding name	-0.014	-0.014	-0.014	-0.014
	(0.011)	(0.011)	(0.011)	(0.011)
AI-Hybrid × Immigrant sounding name	0.023*	0.024*	0.024*	0.025*
	(0.013)	(0.013)	(0.013)	(0.013)
R^2	0.029	0.031	0.030	0.031
Application month–year FE	Yes	Yes	Yes	Yes
Applicant and CV controls	Yes	Yes	Yes	Yes
Candidate–job match controls	Yes	Yes	Yes	Yes
WhatsApp assessment scores	Yes	Yes	Yes	Yes
Score bins	No	Yes	No	Yes
Rank and rank squared	No	No	Yes	Yes
Applications	5434	5434	5434	5434
Job postings	560	560	560	560

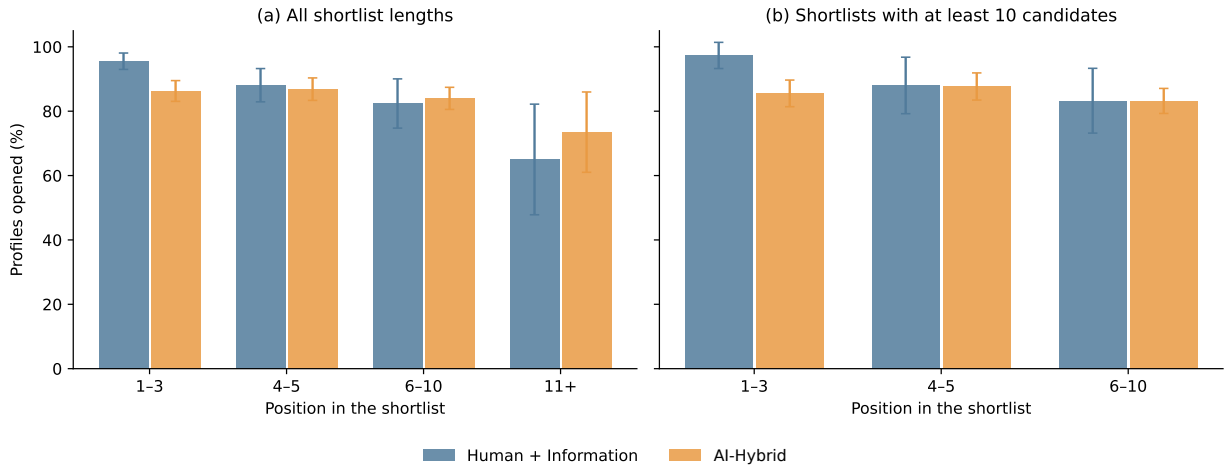
Notes. This table reports estimates of Equation (5) among shortlisted applicants. Panel A uses as dependent variable an indicator for profile opening. Panel B uses as dependent variable an indicator for interview. Panel C ses as dependent variable an indicator for hired. Column (1) reports the baseline specification; column (2) adds AI-score controls, column (3) adds normalized rank and rank squared as controls; column (4) includes both score and rank controls. All columns use the same 5,434 shortlisted applications and 560 postings in July–September 2025. Human ordering follows shortlisting timestamps. AI ordering follows descending scores. Standard errors clustered by employer are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table I10: AI-Hybrid versus Human + Information: employer outcomes controlling for score and ranking (CEM)

	Baseline	+ Score	+ Rank	+ Score and rank
A. Profile opened				
AI-Hybrid	−0.110**	−0.115**	−0.108**	−0.110**
	(0.047)	(0.046)	(0.047)	(0.046)
Immigrant-sounding name	−0.090***	−0.082***	−0.091***	−0.083***
	(0.030)	(0.030)	(0.030)	(0.031)
AI-Hybrid × Immigrant-sounding name	−0.001	0.002	−0.000	0.003
	(0.056)	(0.048)	(0.056)	(0.049)
Control mean	0.937	0.937	0.937	0.937
R^2	0.156	0.195	0.156	0.199
B. Interviewed				
AI-Hybrid	−0.001	−0.002	−0.003	−0.001
	(0.029)	(0.029)	(0.029)	(0.029)
Immigrant-sounding name	−0.032	−0.028	−0.032	−0.028
	(0.023)	(0.023)	(0.023)	(0.023)
AI-Hybrid × Immigrant-sounding name	0.037	0.033	0.037	0.033
	(0.028)	(0.029)	(0.029)	(0.029)
Control mean	0.143	0.143	0.143	0.143
R^2	0.082	0.092	0.082	0.092
C. Hired				
AI-Hybrid	−0.024*	−0.025*	−0.025*	−0.026*
	(0.013)	(0.013)	(0.013)	(0.013)
Immigrant-sounding name	−0.044***	−0.042***	−0.043***	−0.042***
	(0.014)	(0.014)	(0.014)	(0.014)
AI-Hybrid × Immigrant-sounding name	0.051***	0.050***	0.051***	0.050***
	(0.015)	(0.015)	(0.015)	(0.015)
Control mean	0.060	0.060	0.060	0.060
R^2	0.039	0.043	0.041	0.044
Applicant controls	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
CEM weights	Yes	Yes	Yes	Yes
Score bins	No	Yes	No	Yes
Rank and rank squared	No	No	Yes	Yes
Applications	2,929	2,929	2,929	2,929
Job postings	326	326	326	326

Notes. This table reports estimates of Equation (5) among shortlisted applicants. Panel A uses as dependent variable an indicator for profile opening. Panel B uses as dependent variable an indicator for interview. Panel C ses as dependent variable an indicator for hired. Column (1) reports the baseline specification; column (2) adds AI-score controls, column (3) adds normalized rank and rank squared as controls; column (4) includes both score and rank controls. 2,929 shortlisted applications to 326 postings in July–September 2025 CEM sample. Human ordering follows shortlisting timestamps. AI ordering follows descending scores. Standard errors clustered by employer are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure I2: Employer attention along the shortlist



Notes. Percentage of shortlisted profiles opened by employers, by position and screening regime in the July–September sample. Panel (a) includes all shortlist lengths; panel (b) restricts to realized shortlists with at least ten candidates. Each position band includes only profiles occupying those positions. Ordering is reconstructed using shortlisting timestamps and AI scores. Bars report application-level means; intervals are 95 percent confidence intervals clustered by employer.

Table I11: AI-Hybrid versus Human + Information: shortlist size and employer attention

	Shortlisted per posting	Profiles opened per posting
A. All applicants		
AI-Hybrid	5.871***	4.508***
	(0.898)	(0.660)
Human + Information mean	6.205	5.367
B. Applicants with immigrant-sounding names		
AI-Hybrid	1.559***	1.121***
	(0.353)	(0.229)
Human + Information mean	1.843	1.538
Month fixed effects	Yes	Yes
Posting controls	Yes	Yes
Job postings	603	603

Notes. This table reports posting-level regression estimates, using the number of shortlisted applicants as the dependent variable in column (1) and the number of shortlisted applicants whose profiles were opened in column (2). Panel A counts all applicants; Panel B counts applicants with immigrant-sounding names. Each regression uses all 603 postings in the July–September 2025 comparison. Coefficients are expressed in applicants per posting. Standard errors clustered by employer are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

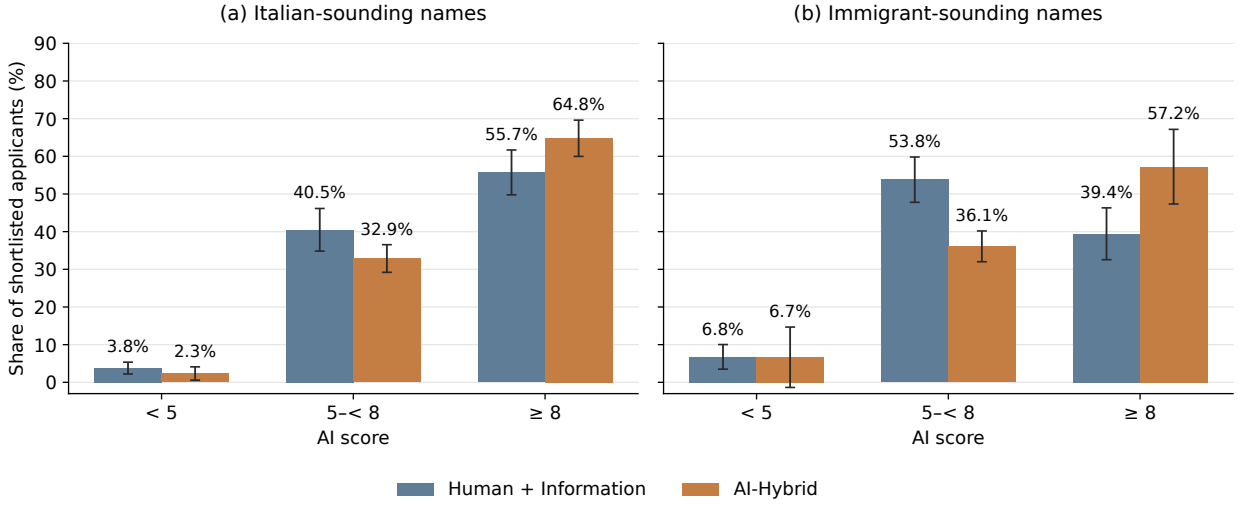
Shortlist size.

Table I12: AI-Hybrid vs Human + Information: shortlist size and expansion by score

A. Shortlist size and name composition				
	Total	Italian-sounding names	Immigrant-sounding names	Immigrant share (pp)
AI-Hybrid	5.871***	3.605***	1.559***	−4.497
	(0.898)	(0.540)	(0.353)	(2.788)
Control mean	6.205	3.933	1.843	32.386
Job postings	603	603	603	579
B. Number shortlisted by name group and AI score				
	Italian-sounding names		Immigrant-sounding names	
	Score 5–< 8	Score ≥ 8	Score 5–< 8	Score ≥ 8
AI-Hybrid	1.211***	2.402***	0.420**	1.055***
	(0.284)	(0.370)	(0.185)	(0.178)
Control mean	1.481	2.038	0.910	0.667
Job postings	603	603	603	603
Month fixed effects	Yes	Yes	Yes	Yes
Posting controls	Yes	Yes	Yes	Yes

Notes. This table reports posting-level regression estimates for July–September 2025. In Panel A, in Column (1) the dependent variable is the total number of shortlisted applicants, in Column (2) is the number with Italian-sounding names, in Column (3) is the number with immigrant-sounding names, and in Column (4) the percentage with immigrant-sounding names among shortlisted applicants. In Panel B, in Column (1) the dependent variable is the number of shortlisted applicants with Italian-sounding names scoring from 5 to below 8. In Column (2) it is the number with Italian-sounding names scoring at least 8. In Column (3) it is the number with immigrant-sounding names scoring from 5 to below 8. In Column (4) it is the number with immigrant-sounding names scoring at least 8. The sample is the pooled sample. Standard errors clustered by employer are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure I3: AI score distribution among shortlisted applicants



Notes. This figure reports the distribution of recorded AI scores among shortlisted applicants in July–September 2025. Panel (a) shows applicants with Italian-sounding names; Panel (b) shows applicants with immigrant-sounding names. The horizontal axis shows AI-score bands. The vertical axis reports the percentage of shortlisted applicants in each band. 95 percent confidence intervals.

Applicant characteristics and job fit.

Table I13: Applicant characteristics and job fit in the complete shortlist

	Age	Female	Total experience	Hospitality experience	Required role experience	Match composite	Occupation match	Distance match	Hours match	Employment gaps
All shortlisted applicants										
AI-Hybrid	2.965***	−0.014	0.758	0.483	0.063	0.018	0.044	−0.045**	−0.005	−0.023
	(0.665)	(0.029)	(0.994)	(0.967)	(0.130)	(0.011)	(0.028)	(0.020)	(0.041)	(0.026)
Control mean	34.205	0.302	8.470	6.605	0.619	0.567	0.515	0.918	0.883	0.555
Job postings	574	580	576	576	577	582	577	580	357	576
R^2	0.210	0.332	0.143	0.155	0.060	0.142	0.194	0.272	0.214	0.088
Italian sounding names										
AI-Hybrid	2.081**	−0.031	0.897	0.479	0.116	0.013	0.020	−0.063***	0.006	0.010
	(0.875)	(0.031)	(0.761)	(0.713)	(0.166)	(0.013)	(0.037)	(0.019)	(0.046)	(0.031)
Control mean	35.854	0.315	9.411	7.544	0.683	0.583	0.548	0.934	0.898	0.585
Job postings	541	554	551	551	549	554	549	551	292	551
R^2	0.175	0.303	0.201	0.214	0.067	0.149	0.175	0.266	0.217	0.127
Immigrant sounding names										
AI-Hybrid	2.632***	0.004	2.289	2.304	0.071	0.051***	0.084*	−0.025	−0.011	−0.025
	(0.862)	(0.039)	(4.183)	(4.181)	(0.197)	(0.016)	(0.043)	(0.029)	(0.080)	(0.044)
Control mean	31.395	0.305	6.142	4.425	0.458	0.555	0.519	0.897	0.859	0.533
Job postings	445	469	460	460	465	469	465	462	153	460
R^2	0.201	0.234	0.080	0.079	0.086	0.123	0.162	0.184	0.340	0.102

Notes. This table reports estimates from regressions at the posting level for the July–September 2025 sample. Column (1) the dependent variable is mean age. In Column (2) the dependent variable is the share of female applicants. In Column (3) the dependent variable is mean total experience. In Column (4) the dependent variable is mean hospitality experience. In Column (5) the dependent variable is mean required-role experience. In Column (6)the dependent variable is the mean composite match index. In Column (7) the dependent variable is mean occupation match. In Column (8) it is mean distance match. In Column (9) it is mean hours match. In Column (10) it is the share of applicants with employment gaps. The sample is the pooled sample of 603 postings, with equal weight given to each posting. Age and experience are measured in years. The other outcomes are shares or indices. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table I14: Applicant characteristics and job fit in complete shortlists in the CEM sample

	Age	Female	Total experience	Hospitality experience	Required role experience	Match composite	Occupation match	Distance match	Hours match	Employment gaps
All shortlisted applicants										
AI-Hybrid	3.598***	−0.036	1.262**	1.193***	0.189	0.009	0.094**	−0.033	−0.075	−0.116**
	(0.858)	(0.063)	(0.528)	(0.420)	(0.161)	(0.013)	(0.047)	(0.033)	(0.068)	(0.045)
p-value	0.000	0.570	0.018	0.005	0.241	0.491	0.047	0.327	0.267	0.011
Control mean	33.450	0.342	7.338	5.256	0.511	0.574	0.483	0.908	0.938	0.611
Job postings	325	330	328	328	332	332	332	330	194	328
R^2	0.315	0.343	0.294	0.348	0.148	0.173	0.263	0.166	0.229	0.189
Italian sounding names										
AI-Hybrid	3.621***	0.002	1.918**	1.366**	0.133	0.017	0.036	−0.073***	−0.042	0.005
	(1.281)	(0.045)	(0.753)	(0.663)	(0.212)	(0.018)	(0.054)	(0.027)	(0.081)	(0.041)
p-value	0.005	0.957	0.012	0.041	0.532	0.357	0.501	0.008	0.601	0.904
Control mean	34.265	0.329	7.548	6.000	0.609	0.583	0.536	0.948	0.926	0.585
Job postings	305	312	311	311	312	312	312	310	154	311
R^2	0.257	0.339	0.237	0.251	0.170	0.193	0.218	0.195	0.209	0.179
Immigrant sounding names										
AI-Hybrid	2.195*	−0.062	0.166	0.756	0.054	0.022	0.093	0.008	−0.182	−0.157**
	(1.215)	(0.065)	(0.859)	(0.656)	(0.226)	(0.020)	(0.066)	(0.049)	(0.178)	(0.064)
p-value	0.073	0.341	0.847	0.251	0.811	0.268	0.160	0.874	0.308	0.015
Control mean	32.364	0.359	6.834	4.475	0.372	0.579	0.543	0.860	0.960	0.635
Job postings	249	263	260	260	263	263	263	260	74	260
R^2	0.379	0.299	0.256	0.297	0.154	0.181	0.247	0.144	0.455	0.200

Notes. This table reports estimates from regressions at the posting level for the July–September 2025 CEM sample, using complete shortlists in both regimes. In Column (1) the dependent variable is mean age. In Column (2) the dependent variable is the share of female applicants. In Column (3) the dependent variable is mean total experience. In Column (4) the dependent variable is mean hospitality experience. In Column (5) the dependent variable is mean required-role experience. In Column (6) the dependent variable is the mean composite match index. In Column (7) the dependent variable is mean occupation match. In Column (8) the dependent variable is mean distance match. In Column (9) the dependent variable is mean hours match. In Column (10) the dependent variable is the share of applicants with employment gaps. The first panel includes all shortlisted applicants. The second panel includes applicants with Italian-sounding names. The third panel includes applicants with immigrant-sounding names. The sample is the CEM sample of 347 postings, using matching weights. The number of postings used varies with group presence and available characteristics. Age and experience are measured in years. The other outcomes are shares or indices. Standard errors clustered at the employer level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table I15: Applicant characteristics and job fit at comparable screening capacity

	Age	Female	Total experience	Hospitality experience	Required role experience	Match composite	Occupation match	Distance match	Hours match	Employment gaps
All shortlisted applicants										
AI-Hybrid	4.752***	−0.029	2.555***	2.287***	0.395*	0.031**	0.125**	−0.019	−0.110	−0.116**
	(0.979)	(0.063)	(0.670)	(0.616)	(0.205)	(0.015)	(0.055)	(0.033)	(0.077)	(0.049)
Control mean	33.450	0.342	7.338	5.256	0.511	0.574	0.483	0.908	0.938	0.611
Job postings	315	327	325	325	329	329	329	326	141	325
R^2	0.300	0.306	0.241	0.242	0.147	0.185	0.224	0.155	0.212	0.204
Italian sounding names										
AI-Hybrid	3.849***	−0.021	2.382***	1.550**	0.306	0.025	0.023	−0.050*	−0.102	−0.006
	(1.407)	(0.048)	(0.838)	(0.738)	(0.293)	(0.021)	(0.058)	(0.028)	(0.098)	(0.054)
Control mean	34.265	0.329	7.548	6.000	0.609	0.583	0.536	0.948	0.926	0.585
Job postings	286	301	300	300	301	301	301	298	111	300
R^2	0.254	0.326	0.227	0.228	0.172	0.211	0.222	0.175	0.212	0.177
Immigrant sounding names										
AI-Hybrid	3.033**	−0.056	1.201	1.593**	0.426	0.035	0.111	−0.014	−0.027	−0.155**
	(1.442)	(0.076)	(0.983)	(0.767)	(0.309)	(0.023)	(0.078)	(0.054)	(0.242)	(0.078)
Control mean	32.364	0.359	6.834	4.475	0.372	0.579	0.543	0.860	0.960	0.635
Job postings	203	220	217	217	220	220	220	216	43	217
R^2	0.399	0.272	0.237	0.283	0.194	0.205	0.234	0.180	0.682	0.205

Notes. This table reports estimates from regressions at the posting level for the July–September 2025 CEM sample with matched size. Column (1) the dependent variable is mean age. In Column (2) the dependent variable is the share of female applicants. In Column (3) the dependent variable is mean total experience. In Column (4) the dependent variable is mean hospitality experience. In Column (5) the dependent variable is mean required-role experience. In Column (6)the dependent variable is the mean composite match index. In Column (7) the dependent variable is mean occupation match. In Column (8) it is mean distance match. In Column (9) it is mean hours match. In Column (10) it is the share of applicants with employment gaps. The sample is the CEM sample of 347 postings, using matching weights. The matched-size set retains the highest-scoring shortlisted applicants up to the rounded mean human shortlist size in the same CEM stratum.Age and experience are measured in years. The other outcomes are shares or indices. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table I16: AI-Hybrid vs Human + Information: soft skills, technical skills, and motivation

	Complete shortlist: pooled			Comparable capacity: CEM matched-size set		
	Soft skills	Technical skills	Motivation	Soft skills	Technical skills	Motivation
All shortlisted applicants						
AI-Hybrid	0.042	−0.076	−0.099	0.183	0.084	0.340*
	(0.091)	(0.193)	(0.137)	(0.128)	(0.328)	(0.203)
Control mean	5.016	4.498	4.927	5.059	4.621	4.745
Job postings	565	565	565	318	318	318
Italian-sounding names						
AI-Hybrid	0.040	0.195	−0.221	0.153	0.456	−0.082
	(0.110)	(0.212)	(0.179)	(0.145)	(0.376)	(0.268)
Control mean	4.994	4.260	4.990	5.092	4.383	4.993
Job postings	540	540	540	293	293	293
Immigrant-sounding names						
AI-Hybrid	0.096	−0.422	0.129	0.423*	0.171	0.600*
	(0.151)	(0.290)	(0.210)	(0.222)	(0.477)	(0.332)
Control mean	4.989	4.644	4.634	5.000	4.711	4.420
Job postings	436	436	436	199	199	199
Posting controls	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table reports estimates from regressions at the posting level for the July–September 2025 CEM sample, using complete shortlists in both regimes. In Column (1) the dependent variable is mean age. In Column (2) the dependent variable is the share of female applicants. In Column (3) the dependent variable is mean total experience. In Column (4) the dependent variable is mean hospitality experience. In Column (5) the dependent variable is mean required-role experience. In Column (6) the dependent variable is the mean composite match index. In Column (7) the dependent variable is mean occupation match. In Column (8) the dependent variable is mean distance match. In Column (9) the dependent variable is mean hours match. In Column (10) the dependent variable is the share of applicants with employment gaps. The first panel includes all shortlisted applicants. The second panel includes applicants with Italian-sounding names. The third panel includes applicants with immigrant-sounding names. The sample is the CEM sample of 347 postings, using matching weights. The number of postings used varies with group presence and available characteristics. Age and experience are measured in years. The other outcomes are shares or indices. Standard errors clustered at the employer level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table I17: Hiring from the complete shortlist, matched-size set and expansion set: CEM

	Complete shortlist	Matched-Size Set	Expansion Set
All shortlisted applicants			
AI-Hybrid	0.081	−0.020	0.119***
	(0.050)	(0.044)	(0.032)
Control mean	0.171	0.171	0.000
Job postings	347	347	347
R^2	0.180	0.155	0.126
Italian sounding names			
AI-Hybrid	0.009	−0.038	0.058**
	(0.045)	(0.041)	(0.023)
Control mean	0.155	0.155	0.000
Job postings	347	347	347
R^2	0.138	0.137	0.081
Immigrant sounding names			
AI-Hybrid	0.057**	0.002	0.060**
	(0.027)	(0.013)	(0.024)
Control mean	0.018	0.018	0.000
Job postings	347	347	347
R^2	0.112	0.100	0.105

Notes. This table reports estimates from regressions at the posting level for July–September 2025. In Column (1) the dependent variable is an indicator for whether the employer hired at least one applicant from the complete shortlist. In Column (2) the dependent variable is an indicator for whether the employer hired at least one applicant from the matched-size set. In Column (3) the dependent variable is an indicator for whether the employer hired at least one applicant from the expansion set. The sample is the CEM sample of 347 postings, using matching weights. All regressions include month fixed effects and predetermined posting controls. Standard errors clustered at the employer level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table I18: Recorded hire counts in the matched-size set and expansion set: CEM

	Complete shortlist	Matched-Size Set	Expansion Set
All shortlisted applicants			
AI-Hybrid	0.091	−0.071	0.161***
	(0.081)	(0.061)	(0.058)
Control mean	0.225	0.225	0.000
Job postings	347	347	347
R^2	0.116	0.118	0.105
Italian sounding names			
AI-Hybrid	0.008	−0.080	0.087**
	(0.068)	(0.055)	(0.042)
Control mean	0.202	0.202	0.000
Job postings	347	347	347
R^2	0.091	0.116	0.070
Immigrant sounding names			
AI-Hybrid	0.061**	0.002	0.059**
	(0.030)	(0.013)	(0.026)
Control mean	0.018	0.018	0.000
Job postings	347	347	347
R^2	0.115	0.100	0.098

Notes. This table reports estimates from regressions at the posting level for July–September 2025. In Column (1) the dependent variable is the number of hires from the complete shortlist. In Column (2) the dependent variable is the number of hires from the matched-size set. In Column (3) the dependent variable is the number of hires from the expansion set. The sample is the CEM sample of 347 postings, using matching weights. All regressions include month fixed effects and predetermined posting controls. Standard errors clustered at the employer level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Employer decisions and hires in the matched-size set and expansion set.

Table I19: Decomposing the change in the ethnic hiring gap into matched-size set and expansion set

	Total hiring	Hiring from matched-size set	Hiring from expansion set
A. Pooled sample: 31,198 applications, 603 postings			
Full controls	0.380**	—	—
	(0.190)		
B. CEM sample: 15,826 applications, 347 postings			
Full controls	0.723**	0.790**	−0.068
	(0.335)	(0.305)	(0.136)
No controls	0.659*	0.772**	−0.113
	(0.341)	(0.314)	(0.135)
Regime-specific control coefficients	0.971**	0.796**	0.175
	(0.403)	(0.360)	(0.197)

Notes. Entries are treatment-by-immigrant-sounding-name coefficients in percentage points, comparing AI-Hybrid with Human + Information. Panel A reproduces the main pooled estimate. Panel B uses CEM weights and the same matched applications for all three outcomes. The full-controls specification includes the deployment application controls and month fixed effects; the final specification additionally interacts these controls and month effects with treatment. In that specification, the reported interaction is evaluated at the covariate reference values rather than averaged over a common covariate distribution. All specifications include treatment, the name indicator, and their interaction. Matched-Size Set membership follows the matching-stratum human-capacity benchmark, with fractional allocation at score ties. Standard errors clustered by employer are in parentheses and treat the estimated matching weights and capacity benchmark as fixed. Components add before rounding. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Where the change in the ethnic hiring gap occurs.

Table I20: Characteristics predicting employer decisions in the equal-capacity matched-size set.

	Profile opened | shortlisted			Interviewed | opened			Hired | opened		
	Human + Information	AI-Hybrid	p-value diff.	Human + Information	AI-Hybrid	p-value diff.	Human + Information	AI-Hybrid	p-value diff.
Non-Italian	-0.040	-0.053*	0.755	-0.117***	-0.040*	0.074	-0.064***	-0.016	0.067
	(0.029)	(0.027)		(0.038)	(0.024)		(0.020)	(0.019)	
Age	-0.004	-0.004	0.988	-0.044	-0.029	0.736	-0.004	-0.033**	0.405
	(0.016)	(0.015)		(0.041)	(0.020)		(0.032)	(0.015)	
Female	0.097**	0.002	0.128	-0.060	-0.023	0.640	-0.049	0.026	0.233
	(0.039)	(0.048)		(0.066)	(0.046)		(0.051)	(0.040)	
Total experience	-0.144	0.081	0.167	0.134	0.010	0.620	0.175	-0.003	0.209
	(0.139)	(0.079)		(0.234)	(0.076)		(0.138)	(0.032)	
Hospitality experience	0.110	-0.079	0.235	-0.094	0.002	0.734	-0.267*	0.024	0.064
	(0.135)	(0.079)		(0.268)	(0.076)		(0.153)	(0.032)	
Experience in required occupation	-0.019	-0.000	0.208	-0.036	0.010	0.245	-0.009	0.005	0.298
	(0.012)	(0.008)		(0.039)	(0.008)		(0.013)	(0.005)	
Occupation match	-0.014	-0.025*	0.651	-0.008	-0.009	0.962	-0.015	-0.003	0.547
	(0.019)	(0.015)		(0.020)	(0.016)		(0.018)	(0.010)	
Distance match	0.016	-0.009	0.211	0.018	-0.021*	0.049	0.010	0.006	0.598
	(0.017)	(0.010)		(0.016)	(0.012)		(0.007)	(0.006)	
Hours match	-0.038**	0.052*	0.005	-0.127	0.016	0.098	-0.169*	-0.013	0.115
	(0.016)	(0.027)		(0.082)	(0.024)		(0.098)	(0.010)	
Employment gaps	0.015	-0.004	0.567	0.042	-0.024	0.308	0.026	-0.049**	0.026
	(0.026)	(0.021)		(0.055)	(0.031)		(0.024)	(0.023)	
Bottom versus top rank	-0.180***	0.050	0.004	-0.142***	-0.188***	0.550	-0.106***	-0.125**	0.790
	(0.049)	(0.063)		(0.050)	(0.066)		(0.037)	(0.057)	
Applications	698	966		594	812		594	812	
Job postings	112	173		103	158		103	158	
R^2	0.578	0.545		0.427	0.458		0.367	0.321	
Posting fixed effects					Yes				
Rank controls					Yes				

I.1.1 Age: hiring from the matched-size set and expansion set

Table I21: Change in the hiring age gradient: matched-size set and expansion set

Specification	Hired	Hired from matched-size set	Hired from expansion set
Pooled: full controls	0.068	—	—
	(0.116)		
CEM: full controls	−0.121	−0.012	−0.109**
	(0.139)	(0.127)	(0.052)
CEM: no controls	−0.168	−0.055	−0.113**
	(0.146)	(0.135)	(0.053)
CEM: regime-specific slopes	0.043	0.175	−0.132***
	(0.152)	(0.145)	(0.049)

Notes. This table reports the coefficient on AI-Hybrid × age from Equation (5) for July–September 2025. Age is measured in ten-year units. In Column (1) the dependent variable is an indicator for whether the applicant was hired. In Column (2) the dependent variable is the hire indicator multiplied by matched-size-set membership. In Column (3) the dependent variable is the hire indicator multiplied by expansion-set membership. The pooled sample contains 26,047 applications to 602 postings. The CEM sample contains 13,204 applications to 346 postings. Standard errors clustered at the employer level are reported in parentheses.

I.1.2 Ethnicity and age on a common sample

Table I22: Common-sample changes in demographic gradients

Specification	Hired	Hired from matched-size set	Hired from expansion set
Immigrant-sounding name	0.950**	1.096***	−0.146
	(0.394)	(0.363)	(0.138)
Age (ten years): Hiring	−0.111	−0.007	−0.104*
	(0.143)	(0.131)	(0.054)

Notes. This table reports the coefficients of Equation (5) on AI-Hybrid × immigrant-sounding name and AI-Hybrid × age from separate regressions for July–September 2025. Age is measured in ten-year units. . In Column (1) the dependent variable is an indicator for whether the applicant was hired. In Column (2) the dependent variable is the hire indicator multiplied by matched-size-set membership. In Column (3) the dependent variable is the hire indicator multiplied by expansion-set membership. CEM sample of 12,724 applications and 346 postings with both non-missing age and non missing immigrant sounding name. Standard errors clustered at the employer level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. p$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

I.1.3 Does flexible age adjustment explain the ethnic result?

Table I23: Ethnic hiring-gap change under alternative age controls

Age specification	Change in hiring gap
Age omitted	0.928**
	(0.394)
Linear age	0.950**
	(0.394)
Linear and quadratic age	0.947**
	(0.395)
Five-year age bands	0.944**
	(0.392)
Age bands interacted with regime	0.939**
	(0.411)

Notes. This table reports the coefficient on AI-Hybrid × immigrant-sounding name from Equation (5) for July–September 2025. The dependent variable is an indicator for whether the applicant was hired. Each row uses a different age specification, retaining the other application controls, CEM weights, and month fixed effects. CEM sample of 12,724 applications, 346 postings. Standard errors clustered at the employer level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$..

I.1.4 Reweighting shortlist size

Table I24: Hiring-gap estimates after reweighting shortlist size

	Change in ethnic hiring gap (pp) (1)	Change in age gradient per ten years (pp) (2)	Effective number of AI-Hybrid postings (3)
Original CEM weights	0.950	-0.111	211.0
Reweighted on shortlist size	1.021	0.043	119.4

Notes. This table reports estimates from Equation (5) for July–September 2025. The dependent variable is an indicator for whether the applicant was hired. In Column (1) the reported coefficient is AI-Hybrid × immigrant-sounding name. In Column (2) the reported coefficient is AI-Hybrid × age. Age is measured in ten-year units. The first row uses the original CEM weights. The second row reweights AI-Hybrid postings toward the shortlist-size distribution under Human + Information. 12,724 applications, 346 postings. Column (3) reports the effective number of AI-Hybrid postings, calculated as $(\sum w)^2 / \sum w^2$ over the 211 AI-Hybrid postings .

J Other deployment comparisons

J.1 Human + Information versus Human only

Table J1: Human + Information versus Human only: Balance

	Pooled				CEM			
	Human only	Human + Information	Std. diff.	p	Human only	Human + Information	Std. diff.	p
Panel A. Job-posting characteristics								
Any language required	0.301	0.311	0.022	0.785	0.288	0.313	0.054	0.599
Italian required	0.006	0.011	0.060	0.428	0.002	0.009	0.069	0.181
Required languages	0.345	0.340	-0.009	0.914	0.323	0.343	0.037	0.717
Maximum distance	28.293	27.505	-0.045	0.712	29.354	26.690	-0.151	0.337
Local residence required	0.014	0.009	-0.054	0.411	0.005	0.004	-0.010	0.783
Minimum experience	1.734	1.646	-0.084	0.245	1.623	1.592	-0.029	0.654
Senior / management role	0.112	0.069	-0.151	0.058	0.056	0.069	0.045	0.577
Customer-facing role	0.450	0.514	0.130	0.082	0.466	0.476	0.020	0.837
Monthly net salary	1728.862	1672.493	-0.126	0.141	1734.118	1719.343	-0.033	0.744
Permanent contract	0.084	0.091	0.026	0.784	0.105	0.116	0.037	0.778
Full-time	0.902	0.911	0.031	0.689	0.911	0.927	0.057	0.538
Days per week	5.622	5.598	-0.043	0.592	5.629	5.547	-0.143	0.169
Split shifts	0.284	0.173	-0.267	0.001	0.305	0.183	-0.291	0.009
Seasonal position	0.324	0.217	-0.241	0.004	0.338	0.189	-0.338	0.002
Number of benefits	2.274	2.176	-0.096	0.339	2.301	2.184	-0.115	0.341
Panel B. Employer characteristics								
Hotel / resort	0.160	0.137	-0.066	0.482	0.135	0.098	-0.103	0.310
Formal restaurant	0.076	0.070	-0.022	0.810	0.068	0.076	0.030	0.797
Staff size	12.492	10.426	-0.243	0.015	12.861	10.579	-0.268	0.050
New employer	0.355	0.549	0.395	< 0.001	0.366	0.528	0.330	0.003
Panel C. Applicant-pool characteristics								
Mean age	36.029	35.395	-0.122	0.132	36.094	35.434	-0.126	0.210
Share female	0.256	0.282	0.148	0.047	0.260	0.252	-0.042	0.626
Mean hospitality experience	5.168	5.079	-0.015	0.846	5.191	5.216	0.004	0.969
Mean experience in required occupation	1.176	0.901	-0.243	< 0.001	1.276	0.968	-0.273	0.005
Job postings	349	350			251	233		

Notes. This table compares job-posting, employer, and applicant-pool characteristics between Human only and Human + Information in May–June 2025. Columns (1) and (2) report unweighted posting-level means; columns (3) and (4) report the standardized difference and the p-value for equality of means. Columns (5)–(8) report the corresponding statistics for matched postings, using CEM weights. CEM p-values come from regressions with standard errors clustered by employer. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J2: Human + Information vs Human only: interviews and hiring

	Pooled		CEM	
	Interviewed	Hired	Interviewed	Hired
Human + Information	0.0010	0.0011	−0.0026	0.0025
	(0.0027)	(0.0017)	(0.0044)	(0.0027)
Immigrant-sounding name	−0.0058**	−0.0014	−0.0128***	−0.0029
	(0.0024)	(0.0020)	(0.0046)	(0.0029)
Human + Information × Immigrant-sounding name	0.0024	0.0005	0.0069	−0.0018
	(0.0027)	(0.0022)	(0.0053)	(0.0030)
Month fixed effects	No	No	No	No
Applicant and CV controls	Yes	Yes	Yes	Yes
Candidate–job match controls	Yes	Yes	Yes	Yes
WhatsApp assessment scores	Yes	Yes	Yes	Yes
Control mean	0.0114	0.0064	0.0124	0.0055
Applications	24533	24533	17117	17117
Job postings	699	699	484	484
R^2	0.0181	0.0123	0.0251	0.0141

Notes. This table reports regression estimates of Equation (5), using an indicator for being interviewed as the dependent variable in columns (1) and (3), and an indicator for being hired in columns (2) and (4). Columns (1) and (2) use all applications in the May–June 2025 comparison sample and give equal weight to each application. Columns (3) and (4) use the CEM matched sample, with each posting's matching weight divided equally among its applications. Human only is the reference regime. Standard errors are clustered at employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J3: Human + Information vs Human only: age, interviews, and hiring

	Pooled		CEM	
	Interviewed	Hired	Interviewed	Hired
Human + Information	0.00883*	0.00265	0.00760	0.00734
	(0.00496)	(0.00359)	(0.00617)	(0.00457)
Age (ten years)	−0.00188*	−0.00189***	−0.00295***	−0.00172**
	(0.00096)	(0.00071)	(0.00111)	(0.00067)
Human + Information × Age (ten years)	−0.00186	−0.00034	−0.00168	−0.00147
	(0.00114)	(0.00080)	(0.00138)	(0.00099)
Month fixed effects	No	No	No	No
Applicant and CV controls	Yes	Yes	Yes	Yes
Candidate–job match controls	Yes	Yes	Yes	Yes
WhatsApp assessment scores	Yes	Yes	Yes	Yes
Immigrant-sounding name control	Yes	Yes	Yes	Yes
Control mean	0.0115	0.0064	0.0121	0.0053
Applications	21,599	21,599	15,097	15,097
Job postings	699	699	484	484
R^2	0.0183	0.0128	0.0271	0.0149

Notes. This table reports estimates of Equation (5) for age, controlling for perceived origin. The dependent variable is an indicator for being interviewed in columns (1) and (3), and for being hired in columns (2) and (4). Age is measured in ten-year units. Columns (1) and (2) use applications with observed age in the May–June 2025 comparison sample and give equal weight to each application. Columns (3) and (4) use the CEM matched sample, with each posting's matching weight divided equally among its applications. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. .

Table J4: Human + Information vs Human only: posting outcomes

	Any interview	Any immigrant-sounding name interview	Any hire	Any immigrant-sounding name hire	Days to shortlist
Human + Information	0.047	0.028	0.035	0.012	0.657**
	(0.035)	(0.025)	(0.030)	(0.021)	(0.277)
Posting controls	Yes	Yes	Yes	Yes	Yes
Month fixed effects	No	No	No	No	No
Control mean	0.229	0.084	0.158	0.058	3.670
Job postings	699	687	699	694	622
R^2	0.083	0.084	0.072	0.077	0.126

Notes. This table reports estimates from regressions at the posting level for May–June 2025. In Column (1) the dependent variable is an indicator for whether the employer interviewed at least one applicant. In Column (2) the dependent variable is an indicator for whether the employer interviewed at least one applicant with an immigrant-sounding name. In Column (3) the dependent variable is an indicator for whether the employer hired at least one applicant. In Column (4) the dependent variable is an indicator for whether the employer hired at least one applicant with an immigrant-sounding name. In Column (5) the dependent variable is mean days to shortlist. The sample is the pooled sample of 699 postings; outcome-specific counts are reported in the table. Standard errors clustered at the employer level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J5: Human + Information vs Human only: employer satisfaction

Outcome	Satisfaction index	Service	Assistance	Recommendation	Reactivate
Human + Information	−0.360	−0.341	−0.398	−0.340	−0.030
	(0.428)	(0.465)	(0.394)	(0.528)	(0.081)
Posting controls	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes
Control mean	7.556	7.017	8.446	7.207	0.769
Observations	258	258	258	258	258
R^2	0.205	0.222	0.153	0.202	0.147

Notes. This table reports regression estimates using as dependent variables the employer satisfaction index in column (1), employers' satisfaction with the service in column (2), satisfaction with the assistance received in column (3), willingness to recommend the platform in column (4), and willingness to reactivate the service in column (5). The outcomes in columns (1)–(4) range from 0 to 10; reactivation is a binary outcome. The satisfaction index is the equally weighted mean of service satisfaction, assistance satisfaction, and willingness to recommend the platform. Feedback is linked to the employer's nearest prior posting within 90 days. Standard errors clustered by employer are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J6: Post-hire outcomes: Human + Information vs Human only

Horizon	30 days	60 days	90 days	180 days
A. Survey-reported separation				
Human + Information	0.165	0.227	0.128	−0.083
	(0.160)	(0.164)	(0.182)	(0.143)
Control mean	0.100	0.200	0.500	0.750
Observed hires	45	45	45	45
B. Subsequent worker application				
Human + Information	0.043	0.007	−0.060	−0.108
	(0.060)	(0.075)	(0.096)	(0.095)
Control mean	0.082	0.164	0.246	0.344
Observed hires	157	157	156	156
C. Employer reposting				
Human + Information	0.038	0.020	0.031	0.008
	(0.064)	(0.067)	(0.080)	(0.101)
Control mean	0.082	0.131	0.164	0.262
Observed hires	157	157	156	155
D. Either administrative signal				
Human + Information	0.071	0.014	−0.042	−0.092
	(0.080)	(0.093)	(0.114)	(0.112)
Control mean	0.148	0.279	0.393	0.525
Observed hires	157	157	156	156

Notes. This table reports regression estimates in which the dependent variable is a post-hire outcomes measured within 30 days in Column (1), 60 days in Column (2), 90 days in Column (3), and 180 days in Column (4). In Panel A, the dependent variable is an indicator for survey-reported separation. In Panel B, the dependent variable is a subsequent worker application. In Panel C, the dependent variable is employer reposting in the same occupation. In Panel D is either administrative. Panel A uses employers' survey responses; Panels B–D use platform records. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J7: Post-hire outcomes by name group: Human + Information vs Human only

Horizon	30 days	60 days	90 days	180 days
A. Survey-reported separation				
Human + Information	0.095	0.164	−0.055	0.010
	(0.228)	(0.249)	(0.220)	(0.187)
Immigrant-sounding name	−0.179	−0.139	−0.459*	0.149
	(0.139)	(0.205)	(0.227)	(0.197)
Human + Information × Immigrant-sounding name	0.159	0.100	0.333	−0.258
	(0.234)	(0.308)	(0.320)	(0.285)
Control mean	0.182	0.273	0.727	0.727
Observations	42	42	42	42
B. Subsequent worker application				
Human + Information	0.007	−0.006	−0.022	−0.072
	(0.090)	(0.096)	(0.107)	(0.118)
Immigrant-sounding name	−0.131**	−0.041	0.123	0.139
	(0.056)	(0.104)	(0.119)	(0.132)
Human + Information × Immigrant-sounding name	0.108	0.037	−0.112	−0.075
	(0.102)	(0.141)	(0.155)	(0.167)
Control mean	0.132	0.184	0.211	0.289
Observations	152	152	151	151
C. Employer reposting				
Human + Information	0.002	0.006	−0.023	−0.005
	(0.062)	(0.067)	(0.081)	(0.116)
Immigrant-sounding name	0.017	0.011	−0.043	0.022
	(0.082)	(0.101)	(0.107)	(0.128)
Human + Information × Immigrant-sounding name	0.087	0.028	0.133	0.024
	(0.111)	(0.125)	(0.145)	(0.168)
Control mean	0.079	0.132	0.184	0.263
Observations	152	152	151	150
D. Either administrative signal				
Human + Information	−0.001	−0.018	−0.063	−0.092
	(0.101)	(0.101)	(0.118)	(0.132)
Immigrant-sounding name	−0.088	−0.004	0.106	0.144
	(0.094)	(0.133)	(0.134)	(0.128)
Human + Information × Immigrant-sounding name	0.191	0.074	0.031	0.010
	(0.138)	(0.166)	(0.171)	(0.165)
Control mean	0.184	0.289	0.368	0.474
Observations	152	152	151	151

Notes. This table reports regression estimates in which the dependent variable is a post-hire outcomes measured within 30 days in Column (1), 60 days in Column (2), 90 days in Column (3), and 180 days in Column (4). In Panel A, the dependent variable is an indicator for survey-reported separation. In Panel B, the dependent variable is a subsequent worker application. In Panel C, the dependent variable is employer reposting in the same occupation. In Panel D is either administrative. Panel A uses employers' survey responses; Panels B–D use platform records. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J8: Human + Information vs Human only: applicant characteristics and job fit at comparable screening capacity

	Age	Female	Match composite	Occupation match	Distance match	Hours match	Hospitality exp. (years)	Minimum exp. met	Employment gaps
Italian sounding names									
Human + Information	−3.438***	−0.107*	−0.011	−0.105	0.043	−0.007	−1.531**	−0.078	−0.011
	(1.256)	(0.062)	(0.024)	(0.067)	(0.037)	(0.103)	(0.673)	(0.064)	(0.066)
Control mean	33.480	0.381	0.597	0.639	0.890	0.842	4.737	0.632	0.592
Immigrant sounding names									
Human + Information	2.618*	0.029	0.060**	0.111	0.032	0.017	1.095	0.139**	0.108
	(1.535)	(0.067)	(0.026)	(0.077)	(0.031)	(0.085)	(0.856)	(0.070)	(0.071)
Control mean	29.932	0.285	0.589	0.591	0.909	0.939	4.083	0.632	0.520
Total matched postings: 484									

Notes. This table reports posting-level regression estimates in the CEM sample. In Column (1), the dependent variable is mean age. In Column (2), the dependent variable is the share of female applicants. In Column (3), the dependent variable is the mean composite match index. In Column (4), the dependent variable is the mean occupation match. In Column (5), the dependent variable is the mean distance match. In Column (6), the dependent variable is the mean hours match. In Column (7), the dependent variable is the mean hospitality experience. In Column (8),the dependent variable is the share of applicants meeting the minimum experience requirement. In Column (9), the dependent variable is the share of applicants with employment gaps. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure J1: Human + Information: shortlisting by recorded AI score and name group

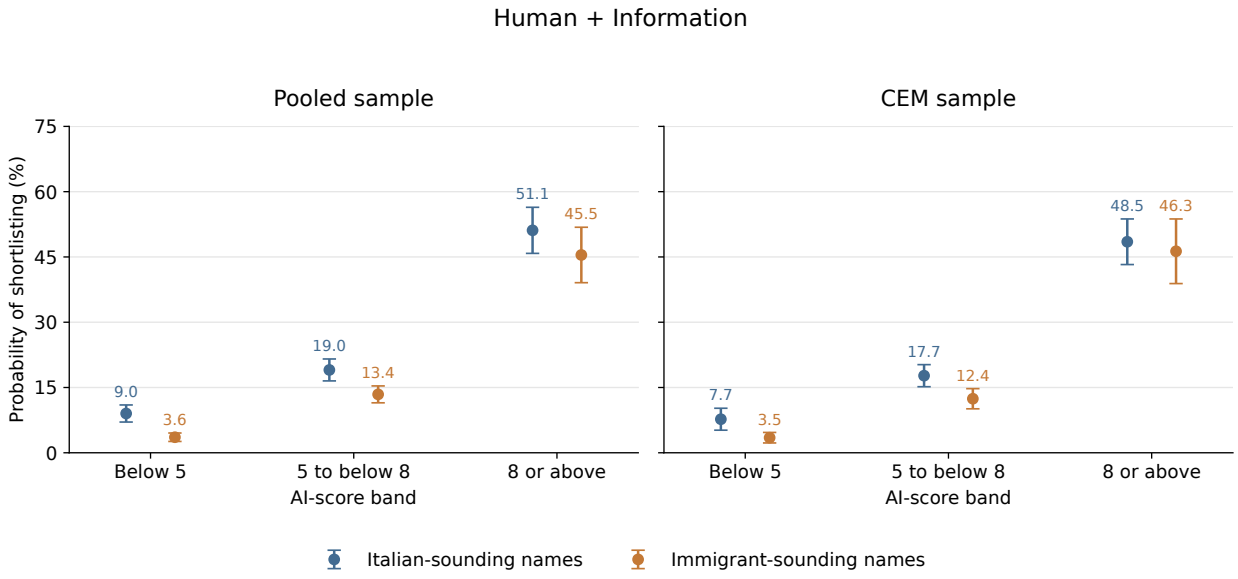
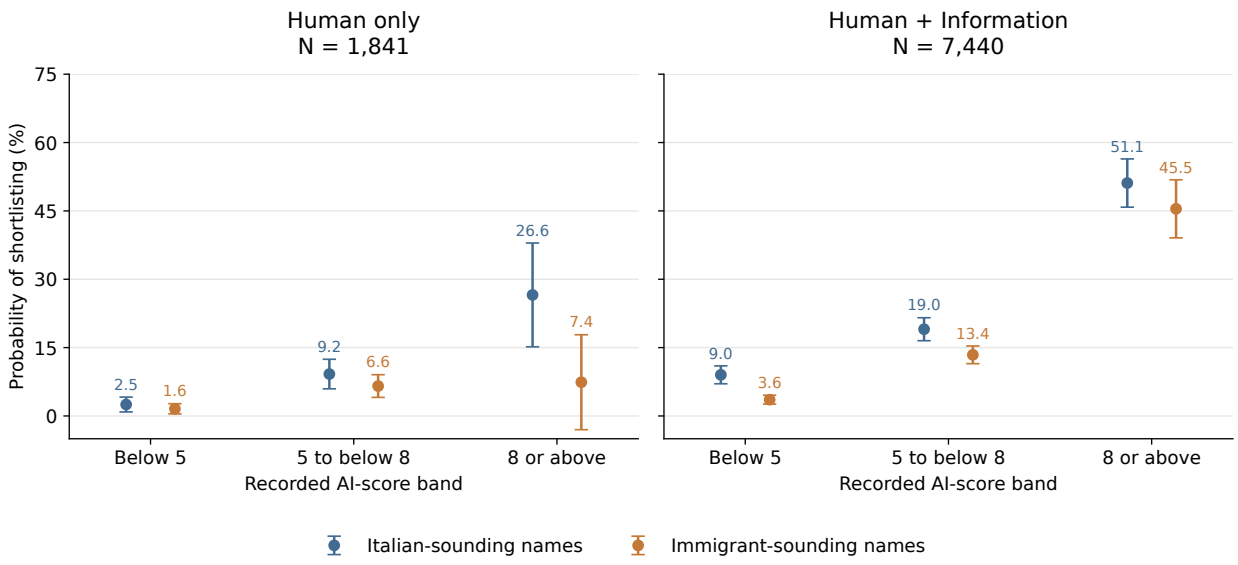


Figure J2: Shortlisting by recorded AI score and name group across human screening regimes



J.2 Fully AI versus AI-Hybrid

Table J9: Fully AI versus AI-Hybrid: Balance

	Pooled				CEM			
	AI-Hybrid	Fully AI	Std. diff.	p	AI-Hybrid	Fully AI	Std. diff.	p
Panel A. Job-posting characteristics								
Any language required	0.313	0.312	-0.004	0.971	0.326	0.312	-0.029	0.834
Italian required	0.004	0.013	0.098	0.369	0.006	0.014	0.087	0.496
Required languages	0.329	0.357	0.052	0.639	0.357	0.348	-0.017	0.898
Maximum distance	24.301	19.179	-0.436	< 0.001	26.049	19.175	-0.586	< 0.001
Local residence required	0.007	0.047	0.251	0.099	0.008	0.051	0.270	0.104
Minimum experience	1.633	1.538	-0.105	0.276	1.624	1.558	-0.074	0.588
Senior / management role	0.106	0.107	0.004	0.965	0.094	0.096	0.006	0.962
Customer-facing role	0.495	0.506	0.022	0.811	0.512	0.500	-0.023	0.850
Monthly net salary	1637.996	1616.672	-0.048	0.632	1615.048	1611.652	-0.008	0.946
Permanent contract	0.123	0.101	-0.071	0.555	0.093	0.096	0.011	0.940
Full-time	0.871	0.828	-0.120	0.280	0.897	0.821	-0.213	0.076
Days per week	5.477	5.562	0.141	0.169	5.429	5.550	0.200	0.108
Split shifts	0.194	0.132	-0.170	0.098	0.192	0.136	-0.152	0.278
Seasonal position	0.061	0.047	-0.060	0.551	0.080	0.038	-0.182	0.290
Number of benefits	2.259	2.331	0.068	0.607	2.358	2.364	0.005	0.971
Panel B. Employer characteristics								
Hotel / resort	0.145	0.100	-0.137	0.260	0.095	0.072	-0.069	0.566
Formal restaurant	0.058	0.080	0.087	0.496	0.054	0.079	0.101	0.474
Staff size	11.114	10.000	-0.170	0.225	10.377	10.120	-0.039	0.823
New employer	0.227	0.213	-0.034	0.758	0.192	0.192	0.000	1.000
Panel C. Applicant-pool characteristics								
Mean age	36.098	35.114	-0.244	0.014	36.299	35.035	-0.313	0.013
Share female	0.300	0.334	0.203	0.052	0.306	0.325	0.109	0.350
Mean hospitality experience	6.385	5.855	-0.075	0.447	5.535	5.970	0.062	0.394
Mean experience in required occupation	0.468	0.553	0.089	0.402	0.484	0.568	0.088	0.462
Job postings	295	169			244	156		

Notes. This table compares job-posting, employer, and applicant-pool characteristics between AI-Hybrid and Fully AI in September–October 2025. Columns (1) and (2) report unweighted posting-level means; columns (3) and (4) report the standardized difference and the p-value for equality of means. Columns (5)–(8) report the corresponding statistics for matched postings, using CEM weights. CEM p-values come from regressions with standard errors clustered by employer. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J10: Fully AI vs AI-Hybrid: interviews and hiring

	Pooled		CEM	
	Interviewed	Hired	Interviewed	Hired
Fully AI	0.00483	0.00197	0.00580	0.00206
	(0.00560)	(0.00178)	(0.00626)	(0.00173)
Immigrant-sounding name	0.00367	0.00211	0.00715*	0.00341*
	(0.00258)	(0.00142)	(0.00369)	(0.00176)
Fully AI × Immigrant-sounding name	−0.00454	−0.00174	−0.00735	−0.00186
	(0.00375)	(0.00194)	(0.00477)	(0.00205)
Month fixed effects	Yes	Yes	Yes	Yes
Applicant and CV controls	Yes	Yes	Yes	Yes
Candidate–job match controls	Yes	Yes	Yes	Yes
WhatsApp assessment scores	Yes	Yes	Yes	Yes
Control mean	0.0213	0.0061	0.0165	0.0033
Applications	28,927	28,927	24,557	24,557
Job postings	464	464	400	400
R^2	0.0353	0.0141	0.0394	0.0146

Notes. This table reports regression estimates of Equation (5), using an indicator for being interviewed as the dependent variable in columns (1) and (3), and an indicator for being hired in columns (2) and (4). Columns (1) and (2) use all applications in the September–October 2025 comparison sample and give equal weight to each application. Columns (3) and (4) use the CEM matched sample, with each application receiving its posting's matching weight. AI-Hybrid is the reference regime. Standard errors are clustered at employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J11: Fully AI vs AI-Hybrid: age, interviews, and hiring

	Pooled		CEM	
	Interviewed	Hired	Interviewed	Hired
Fully AI	0.00577	0.00286	0.00193	0.00184
	(0.00924)	(0.00333)	(0.00894)	(0.00319)
Age (ten years)	−0.00323***	−0.00126*	−0.00353***	−0.00136**
	(0.00123)	(0.00065)	(0.00132)	(0.00054)
Fully AI × Age (ten years)	−0.00051	−0.00050	0.00049	−0.00015
	(0.00192)	(0.00078)	(0.00190)	(0.00069)
Month fixed effects	Yes	Yes	Yes	Yes
Applicant and CV controls	Yes	Yes	Yes	Yes
Candidate–job match controls	Yes	Yes	Yes	Yes
WhatsApp assessment scores	Yes	Yes	Yes	Yes
Immigrant-sounding name control	Yes	Yes	Yes	Yes
Control mean	0.0199	0.0059	0.0143	0.0029
Applications	23,266	23,266	19,807	19,807
Job postings	463	463	399	399
R^2	0.0350	0.0144	0.0411	0.0148

Notes. This table reports estimates of Equation (5) for age, controlling for perceived origin. The dependent variable is an indicator for being interviewed in columns (1) and (3), and for being hired in columns (2) and (4). Age is measured in ten-year units. Columns (1) and (2) use applications with observed age in the September–October 2025 comparison sample and give equal weight to each application. Columns (3) and (4) use the CEM matched sample, with each application receiving its posting's matching weight. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. .

Table J12: Fully AI vs AI-Hybrid: posting outcomes

Outcome	Any interview	Any immigrant-sounding name interview	Any hire	Any immigrant-sounding name hire	Days to shortlist
Fully AI	0.06089	−0.01765	0.04260	−0.00885	0.28847
	(0.06386)	(0.05496)	(0.05917)	(0.03718)	(0.19253)
Posting controls	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes
Control mean	0.434	0.237	0.275	0.113	2.339
Job postings	464	437	464	444	453
R^2	0.148	0.098	0.130	0.108	0.137

Notes. This table reports estimates from regressions at the posting level for September–October 2025. In Column (1) the dependent variable is an indicator for whether the employer interviewed at least one applicant. In Column (2) the dependent variable is an indicator for whether the employer interviewed at least one applicant with an immigrant-sounding name. In Column (3) the dependent variable is an indicator for whether the employer hired at least one applicant. In Column (4) the dependent variable is an indicator for whether the employer hired at least one applicant with an immigrant-sounding name. In Column (5) the dependent variable is mean days to shortlist. The sample is the pooled sample of 464 postings; outcome-specific counts are reported in the table. Application-level delays exceeding 30 calendar days are treated as missing in both regimes. Standard errors clustered at the employer level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J13: Fully AI vs AI-Hybrid: employer satisfaction

Outcome	Satisfaction index	Service	Assistance	Recommendation	Reactivate
Fully AI	0.25883	0.25299	0.28072	0.24279	0.04520
	(0.36541)	(0.40826)	(0.35594)	(0.41548)	(0.07202)
Posting controls	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes
Control mean	7.599	7.293	8.106	7.398	0.805
Observations	197	197	197	197	197
R^2	0.229	0.195	0.252	0.235	0.231

Notes. This table reports regression estimates using as dependent variables the employer satisfaction index in column (1), employers' satisfaction with the service in column (2), satisfaction with the assistance received in column (3), willingness to recommend the platform in column (4), and willingness to reactivate the service in column (5). The outcomes in columns (1)–(4) range from 0 to 10; reactivation is a binary outcome. The satisfaction index is the equally weighted mean of service satisfaction, assistance satisfaction, and willingness to recommend the platform. Feedback is linked to the employer's nearest prior posting within 90 days. Standard errors clustered by employer are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J14: Post-hire outcomes: Fully AI vs AI-Hybrid

Horizon	30 days	60 days	90 days	180 days
A. Survey-reported separation				
Fully AI	0.14286	0.04762	0.04762	0.20238
	(0.26981)	(0.24911)	(0.24911)	(0.22043)
Control mean	0.238	0.429	0.429	0.524
Observed hires	36	36	36	36
B. Subsequent worker application				
Fully AI	−0.08090	−0.06344	−0.02209	0.14675
	(0.05525)	(0.06712)	(0.06833)	(0.11932)
Control mean	0.112	0.142	0.171	0.292
Observed hires	183	182	179	98
C. Employer reposting				
Fully AI	0.12534	0.07558	0.03787	0.12051
	(0.11485)	(0.11796)	(0.11857)	(0.09270)
Control mean	0.243	0.308	0.393	0.658
Observed hires	183	183	183	119
D. Either administrative signal				
Fully AI	0.04259	0.01505	−0.00044	0.09857
	(0.11642)	(0.11459)	(0.10856)	(0.06682)
Control mean	0.336	0.421	0.523	0.762
Observed hires	183	183	183	131

Notes. This table reports regression estimates in which the dependent variable is a post-hire outcomes measured within 30 days in Column (1), 60 days in Column (2), 90 days in Column (3), and 180 days in Column (4). In Panel A, the dependent variable is an indicator for survey-reported separation. In Panel B, the dependent variable is a subsequent worker application. In Panel C, the dependent variable is employer reposting in the same occupation. In Panel D is either administrative. Panel A uses employers' survey responses; Panels B–D use platform records. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J15: Post-hire outcomes by name group: Fully AI vs AI-Hybrid

Horizon	30 days	60 days	90 days	180 days
A. Survey-reported separation				
Fully AI	0.15027	0.02216	0.02216	0.26640
	(0.30912)	(0.27848)	(0.27848)	(0.22636)
Immigrant-sounding name	−0.11237	−0.28059	−0.28059	−0.26263
	(0.21106)	(0.25323)	(0.25323)	(0.26114)
Fully AI × Immigrant-sounding name	0.09508	0.17332	0.17332	−0.13675
	(0.48328)	(0.50155)	(0.50155)	(0.48189)
Control mean	0.273	0.545	0.545	0.636
Observations	31	31	31	31
B. Subsequent worker application				
Fully AI	−0.05141	−0.07427	−0.04580	0.09706
	(0.07132)	(0.08391)	(0.08747)	(0.13985)
Immigrant-sounding name	−0.00821	−0.06210	−0.10121	−0.10028
	(0.06842)	(0.07146)	(0.07488)	(0.10424)
Fully AI × Immigrant-sounding name	−0.09018	0.03107	0.08266	0.20893
	(0.09656)	(0.10630)	(0.12059)	(0.16528)
Control mean	0.121	0.175	0.214	0.308
Observations	160	159	157	90
C. Employer reposting				
Fully AI	0.23199**	0.14046	0.05207	0.10642
	(0.11216)	(0.12065)	(0.12425)	(0.12186)
Immigrant-sounding name	0.07991	0.05040	0.00346	−0.08076
	(0.08964)	(0.09319)	(0.10742)	(0.12225)
Fully AI × Immigrant-sounding name	−0.21417*	−0.14303	0.05245	0.07161
	(0.11389)	(0.13260)	(0.16979)	(0.17468)
Control mean	0.190	0.276	0.379	0.667
Observations	160	160	160	105
D. Either administrative signal				
Fully AI	0.12036	0.02046	−0.02671	0.05752
	(0.11957)	(0.12440)	(0.11725)	(0.08752)
Immigrant-sounding name	0.04380	−0.01991	−0.08491	−0.11085
	(0.10548)	(0.11172)	(0.10996)	(0.09864)
Fully AI × Immigrant-sounding name	−0.20551	−0.03477	0.15070	0.14195
	(0.13058)	(0.15156)	(0.16497)	(0.11236)
Control mean	0.310	0.431	0.552	0.783
Observations	160	160	160	117

Notes. This table reports regression estimates in which the dependent variable is a post-hire outcomes measured within 30 days in Column (1), 60 days in Column (2), 90 days in Column (3), and 180 days in Column (4). In Panel A, the dependent variable is an indicator for survey-reported separation. In Panel B, the dependent variable is a subsequent worker application. In Panel C, the dependent variable is employer reposting in the same occupation. In Panel D is either administrative. Panel A uses employers' survey responses; Panels B–D use platform records. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J16: Fully AI vs AI-Hybrid: shortlist expansion and employer attention

	Pooled			CEM		
	Shortlisted	Profiles opened	Share opened (%)	Shortlisted	Profiles opened	Share opened (%)
Fully AI	13.000***	5.991***	−14.095***	12.634***	5.700***	−15.577***
	(2.000)	(1.443)	(3.629)	(2.006)	(1.381)	(4.023)
Control mean	12.841	10.620	86.411	12.402	10.197	86.827
Posting controls	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Job postings	464	464	453	400	400	391

Notes. This table reports posting-level regression estimates for September–October 2025. In Column (1) the dependent variable is the number of shortlisted applicants. In Column (2) the dependent variable is the number of shortlisted applicants whose profiles were opened. In Column (3) the dependent variable is the percentage of shortlisted profiles opened. In Column (4) the dependent variable is the number of shortlisted applicants. In Column (5) the dependent variable is the number of shortlisted applicants whose profiles were opened. In Column (6) the dependent variable is the percentage of shortlisted profiles opened. Columns (1)–(3) use the pooled sample of 464 postings, giving equal weight to each posting. Columns (4)–(6) use the CEM sample of 400 postings, using matching weights. Columns (3) and (6) use postings with nonempty shortlists. Count coefficients are expressed in applicants per posting; share coefficients are in percentage points. Standard errors clustered at the employer level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J17: Selection within vacancies: five-component match index and its components

	Shortlisted means		Shortlisted minus non-shortlisted			
	AI-Hybrid	Fully AI	AI-Hybrid	Fully AI	Difference	p-value
A. Pooled						
Composite match	0.633	0.619	0.088***	0.094***	0.006	0.725
			(0.012)	(0.013)	(0.018)	
Distance match	0.934	0.971	0.148***	0.367***	0.220***	< 0.001
			(0.023)	(0.041)	(0.047)	
Occupation match	0.482	0.387	0.119***	−0.015	−0.134***	0.001
			(0.032)	(0.023)	(0.040)	
Experience match	0.111	0.067	0.059***	−0.006	−0.065***	0.009
			(0.021)	(0.011)	(0.025)	
Language match	0.625	0.708	0.069*	0.039	−0.030	0.599
			(0.039)	(0.041)	(0.057)	
Sector match	0.855	0.854	0.021	0.014	−0.008	0.768
			(0.022)	(0.015)	(0.027)	
B. CEM						
Composite match	0.625	0.624	0.077***	0.100***	0.023	0.256
			(0.015)	(0.013)	(0.020)	
Distance match	0.915	0.974	0.153***	0.387***	0.234***	< 0.001
			(0.041)	(0.041)	(0.058)	
Occupation match	0.547	0.389	0.128***	−0.016	−0.143***	0.006
			(0.044)	(0.024)	(0.051)	
Experience match	0.111	0.071	0.061**	−0.003	−0.064*	0.062
			(0.031)	(0.012)	(0.034)	
Language match	0.553	0.718	0.019	0.052	0.034	0.660
			(0.065)	(0.041)	(0.077)	
Sector match	0.845	0.855	−0.014	0.012	0.026	0.438
			(0.030)	(0.016)	(0.034)	

Notes. This table reports estimates for applicants with immigrant-sounding names and recorded AI scores from 5 to below 8 in September–October 2025. In Column (1), we report the mean among shortlisted applicants under AI-Hybrid. In Column (2), we report the mean among shortlisted applicants under Fully AI. In Column (3), we report the difference between shortlisted and non-shortlisted applicants within the same vacancy under AI-Hybrid. In Column (4), we report this difference under Fully AI. Column (5) reports Column (4) minus Column (3). Column (6) reports the p-value for equality of the two differences. Panel A uses the pooled sample. Panel B uses the CEM sample. Standard errors are clustered at the employer level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J18: Fully AI vs AI-Hybrid: applicant characteristics and job fit at comparable screening capacity

	Age	Female	Match composite	Occupation match	Distance match	Hours match	Hospitality exp. (years)	Minimum exp. met	Employment gaps
Italian sounding names									
Fully AI	−0.269	−0.048*	−0.019	−0.042	−0.047	0.057	1.489	−0.011	−0.030
	(1.047)	(0.029)	(0.014)	(0.040)	(0.032)	(0.054)	(1.872)	(0.029)	(0.032)
Control mean	37.906	0.308	0.599	0.609	0.848	0.836	6.822	0.734	0.626
Immigrant sounding names									
Fully AI	−1.193	−0.001	−0.042**	−0.127***	−0.058*	−0.039	−1.034*	−0.104***	−0.020
	(1.153)	(0.044)	(0.017)	(0.045)	(0.034)	(0.094)	(0.567)	(0.038)	(0.042)
Control mean	32.778	0.330	0.606	0.604	0.934	0.836	5.786	0.733	0.532
Total matched postings: 399									

Notes. This table reports estimates from regressions at the posting level for the September–October 2025 CEM sample at comparable screening capacity. In Column (1) the dependent variable is mean age. In Column (2) the dependent variable is the share of female applicants. In Column (3) the dependent variable is the mean composite match index. In Column (4) the dependent variable is mean occupation match. In Column (5) the dependent variable is mean distance match. In Column (6) the dependent variable is mean hours match. In Column (7) the dependent variable is mean hospitality experience. In Column (8) the dependent variable is the share of applicants meeting the minimum experience requirement. In Column (9) the dependent variable is the share of applicants with employment gaps. The sample is the CEM sample of 399 postings, using matching weights. The Fully AI matched-size set retains the highest-scoring shortlisted applicants up to the rounded mean AI-Hybrid shortlist size in the same CEM stratum. Shorter lists enter in full, and cutoff ties receive fractional membership. AI-Hybrid postings contribute their complete shortlist. Age and hospitality experience are measured in years. The other outcomes are shares or indices. Standard errors clustered at the employer level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table J19: Fully AI vs AI-Hybrid: robustness to matching on distance requirements

	Interviewed	Hired
Fully AI	0.0044	0.0006
	(0.0070)	(0.0026)
Immigrant-sounding name	0.0050	0.0020
	(0.0042)	(0.0021)
Fully AI × immigrant-sounding name	−0.0058	0.0004
	(0.0053)	(0.0029)
Applicant and CV controls	Yes	Yes
Candidate–job match controls	Yes	Yes
WhatsApp assessment scores	Yes	Yes
Month fixed effects	Yes	Yes
Control mean	0.0179	0.0046
Applications	17,461	17,461
Job postings	269	269
R^2	0.0400	0.0188

Notes. This table reports regression estimates of Equation (5) for September–October 2025. In Column (1) the dependent variable is an indicator for whether the applicant was interviewed. In Column (2) the dependent variable is an indicator for whether the applicant was hired. The CEM sample contains 17,461 applications to 269 postings. Matching additionally uses the maximum applicant distance specified by the employer: up to 10 km, over 10 to 25 km, over 25 to 50 km, and over 50 km. Missing distances form a separate category. The other matching dimensions are month, occupation, salary band, and new-employer status. Each application receives its posting's matching weight. Regressions retain the application controls and month fixed effects from Equation (5). Standard errors clustered at the employer level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

K A Framework for Screening Signals and Employer Response

For our framework, we build on the sequential approach of Harrington and Shaffer (2026): downstream decisions are a function of the assessments made by earlier actors. In our setting screening determines which applicants an employer can even consider. The framework distinguishes between applicant fit, screening assessments and employer choice. This framework allows us to structure the empirical comparisons. Neither screening technologies is unambiguously better.

K.1 Delegation and information

To reduce searching costs, employers delegate initial screening to a platform. The platform gathers information on both applicants and vacancies and then forwards a shortlisted set

of candidates to the employer. The employer then has complete control over the process of further consideration, interviews and the actual hiring (including the option of not filling the vacancy at all). We take the applicant pool as given, the framework doesn't model any equilibrium effects in the number of applications or the number of vacancies created. Delegation requires employers to communicate their requirements. Employers can record some criteria, such as experience, language skills, and availability, explicitly. Some other dimensions, among which group preferences, could be harder to describe and learned through repeated interactions between reviewers and employers. Screeners interpret the available information. Human reviewers may also impose their own biases. A large language model (LLM) could process information differently and at a different cost. Also it could introduce or maintain group disparities.

K.2 Applicant fit and screening assessments

Let ' s write the applicant -vacancy fit as:

$$\theta_{ij} = g(X_{ij}) + u_{ij}, \tag{9}$$

where i is the applicant and j is the job posting. u_{ij} captures the additional employer-specific fit which is not fully captured by the recorded information X_{ij}.

Then, let s^r_{ij} denote the screening assessment produced by human reviewers ($r = H$) or AI ($r = AI$). s^r_{ij} does not need to be an explicitly recorded numerical score.

$$s^r_{ij} = f_r(X_{ij}) + \lambda_r u_{ij} + b_{rj} G_i + \varepsilon^r_{ij}, \qquad r \in \{H, AI\}, \tag{10}$$

where G_i is an indicator for whether the applicant has an immigrant-sounding name. The function f_r describes the regime-specific evaluation of recorded characteristics and allows for nonlinearities and interactions. The term $\lambda_r u_{ij} + \varepsilon^r_{ij}$ represents an imperfect assessment of additional fit, with $\lambda_r \in [0, 1]$ and assessment noise ε^r_{ij}. The screener need not observe u_{ij} without error. We assume neither $f_H = g$ nor $f_{AI} = g$, and do not impose $\lambda_H > \lambda_{AI}$.

The parameter b_{rj} is a weight that adjusts the screening score based on groups. If $b_{rj} < 0$ and everything else equal applicants with immigrant sounding names will have lower scores. In human screening b_{rj} depends on the reviewer identity $\rho(j)$ and on employer identity $e(j)$:

$$b_{Hj} = b_{H,\rho(j),e(j)}.$$

K.3 Shortlist and employer consideration

Screening assessments determine which applicants are sent to the employer. Let $\mathscr{L}_j^r$ be the shortlist under screening rule r. A change in the rule will affect both who is selected and how many applicants are sent forward.

Let $\mathscr{C}_j^r \subseteq \mathscr{L}_j^r$ denote the subset that employer consider. $\mathscr{C}_j^r$ may depend on: (i) ranking (ii) info (iii) size.

K.4 Employer assessment and choice

Let's indicate with $\tilde{\pi}_{e,ij}$ the employer's assessment of the fit between the applicant and the vacancy. Therefore, the value of hiring the candidate is:

$$V_{e,ij} = \tilde{\pi}_{e,ij} - t_e G_i, \tag{11}$$

where t_e is group specific preference.

For a vacancy with one post the employer will draw from the candidates under consideration and the option of leaving the vacancy unfilled:

$$i_j^r \in \arg\max_{i \in \mathscr{C}_j^r \cup \{0\}} V_{e,ij}, \qquad V_{e,0j} = V_{e0}. \tag{12}$$

Our analysis doesn't separately identify employers' assessments of fit from group preferences. The alignment with employer choice therefore doesn't necessarily mean that the applicant is more productive.

K.5 Conditional implications

(i) Changing a group adjustment is different from changing a screening rule.:

Moving from human to AI screening can change several aspects (i) how applicant characteristics are evaluated (ii) how additional fit is assessed, and (iii) the adjustment associated with group membership. A less adverse group adjustment does not imply a better representation.

Changing the screening rule changes more than the group adjustment. Moving from human to AI screening can change how applicant characteristics are evaluated, how additional fit is assessed, and the adjustment associated with group membership. A less

adverse group adjustment can therefore be counteracted by changes in the assessment of other characteristics. The change in group representation on the shortlist does not, by itself, identify a change in b_{rj}.

Useful information and adverse group adjustments can coexist. For two applicants in the same group within a vacancy:

$$s^r_{ij} - s^r_{kj} = f_r(X_{ij}) - f_r(X_{kj}) + \lambda_r(u_{ij} - u_{kj}).$$

The group adjustment cancels because both applicants belong to the same group. Holding f_r fixed, increasing λ_r improves the relative score of the applicant with higher additional fit.

As a simple benchmark without assessment noise, suppose $f_H = f_{AI} = g$ and $\lambda_H = 1 > \lambda_{AI}$. Human scores then rank applicants within the same group by actual fit, whereas AI scores may reverse that ranking. If $b_{Hj} < 0$, the human rule incorporates useful information while also penalizing applicants with immigrant-sounding names relative to other applicants. This does not establish that human screening is better overall: evaluation functions and assessment noise may differ across technologies.

Broader access can change who gets the job without in any way changing the fact that a job is being offered. An adverse screening adjustment could result in the exclusion of a candidate that the employer would have preferred to be among those forwarded. A change in screening could make that candidate available for consideration and so change the identity or group of the hire, with the vacancy still being filled under both regimes. Improved access doesn't necessarily lead to hiring if the candidate is not considered or if the employer prefers a different candidate. The group hiring outcome will therefore be a function of both the screening rule and the employer's choice.

The same argument can be made for age, using an age-group indicator or a more flexible age-dependent assessment. Different downstream patterns for ethnicity and age are consistent with differences in employer responses, but may also reflect the characteristics and job fit of newly selected candidates. They do not establish that reviewers correctly anticipate employer preferences for one demographic dimension but not another.

Capacity expansion has a monotonic benchmark (i.e. it's not an unconditionally beneficial thing to do). Hold the screening ranking fixed and increase the shortlist from K to K' > K so that $L^r_j(K)$ is a subset of $L^r_j(K')$. If all the candidates are available at zero cost and

employer valuations are unchanged then

$$\max\left\{V_{e0}, \max_{i\in\mathscr{L}_j^r(K')} V_{e,ij}\right\} \geq \max\left\{V_{e0}, \max_{i\in\mathscr{L}_j^r(K)} V_{e,ij}\right\}. \tag{13}$$

Expansion can make a preferred candidate available without any increase in hiring if the employer would have filled the post anyway. It needn't have any effect on the group of candidate that would have been hired in that particular direction.

Costly attention on its own doesn't mean that expansion makes the employer worse off. If the employer can just go on to ignore the additional candidates at no cost and continue with the previous search strategy then that strategy is still viable. The only way in which it could be detrimental would be if some other change occurred (e.g. higher navigation costs, lower visibility of the previous set of candidates, an attention allocation process which doesn't preserve the old consideration set). Human-AI comparisons don't necessarily need to meet this benchmark as rankings, capacity and presentation can all change together. Faster processing is another empirical margin of course and the impact of that will depend on the assumptions we make about the timing of the search which are not part of this analysis.

K.6 Mapping to the empirical analysis

The same applicant comparison fixes the pool and shortlist size and then compares who the two rules select. The characteristics of selection, the conditional group gaps and their dispersion describe the combined rule change. They don't separately recover the information advantage, the group adjustment or the processing shortcuts.

The deployment comparison changes the screening process (including the allocation of authority between reviewer and model) and as such can change the characteristics and job fit of the selected applicants, the size of the shortlist and the timing of the shortlist. Changes in hiring gaps without a detectable change in aggregate hiring are consistent with a reallocation of opportunities, though an aggregate null does not necessarily mean that every vacancy would be filled under both regimes.

The mechanism analysis looks at the applicant characteristics and job fit in the comparable capacity, the ranking and employer activity and the location of the hires within both the matched size set and the expansion set of realised shortlists. Hires from the expansion set indicate that the employer did in fact select an applicant outside of the highest ranked matched size set. It doesn't tell us what they would have done had the shortlist been smaller or whether the human reviewers would have selected those applicants in preference.

Employer satisfaction will be a function of both the recruitment outcomes and the cost of using the service (including the effort of assessing the candidates and the speed of the process). Screening can therefore have an impact on satisfaction even if the actual hiring is unchanged. Satisfaction is an assessment of the service rather than a measure of the employer value or productivity in Equation (11). It is something that is observed in the respondents and the composition of the respondents may differ across regimes. Retention is a different margin entirely: it's about the durability of the matches that are made.