Jobseeker Engagement and Selection on Matching Platforms: Field and Lab Experiments on AI Disclosure

Abstract

Artificial intelligence is becoming more deeply embedded in labor-market matching platforms. How well such a platform works depends not only on whom it recommends, but also on who chooses to engage with those recommendations. We study this participation margin in a field experiment involving 4,465 jobseekers, randomly varying whether the same candidate–job match was attributed to an AI matching algorithm, a human employer, or no source. AI disclosure reduced engagement by about 26 percent relative to both explicit human matching and the no-disclosure control. More importantly, the decline was concentrated among better-matched jobseekers, changing the match-quality composition of the engaged pool. We investigate several possible explanations using evidence from the field experiment and a companion lab experiment. Across the two experiments, the evidence is consistent with several interpretations and suggests that disclosing AI use changes jobseekers' expectations about how they are likely to fare within the matching process. We discuss implications for the design of matching platforms, the use and disclosure of AI in matching, and why better prediction need not translate into better realized matching.

Keywords: platform strategy; matching markets; matching-system design; endogenous selection; artificial intelligence; disclosure; labor-market matching; field experiments

1. INTRODUCTION

Artificial intelligence is rapidly becoming embedded in labor-market matching platforms. An estimated 90 percent of large U.S. firms use some form of digital automation in hiring (Fuller and Raman, 2021), with systems ranging from keyword filters to semantic embeddings and generative AI. Research has accordingly focused on improving the matching technology: better algorithms, richer data, and more accurate predictions of candidate–job match quality (e.g., Le Barbanchon et al., 2023). Much of this work, however, treats applicants as *inputs* to the system rather than as decision-makers who respond to it, even as algorithms increasingly determine who is considered at all (Fuller and Raman, 2021). A long labor-market literature likewise studies how workers and vacancies are paired once both sides are present (Mortensen, 1986; Pissarides, 2000). But whether jobseekers choose to engage is itself part of the matching problem. The problem is more general to matching markets: the mechanism used to identify, recommend, or screen potential matches can affect whether participants enter the process at all, making the pool available for subsequent matching endogenous to the matching system.

A labor-market matching platform must do more than pair jobseekers with employers. It must first

create the pool of jobseekers willing to engage with a given position. Before any application or screen, a jobseeker decides whether a posting is worth pursuing at all. We call that choice *engagement*. Everything beyond it is costly: reading the posting closely, tailoring an application, attending an interview. None of it happens if the candidate does not engage. A matching system's predictions therefore matter only if jobseekers act on them, and *which* jobseekers act determines the pool from which any later match to a job can be made. Who enters depends partly on what the platform does and what it discloses about how it matches. Candidates who do not engage at this stage leave no later-stage record, which makes this margin unusually hard to observe and impossible for later screening or selection to recover.

This paper asks whether disclosing AI matching changes not only how many jobseekers engage, but how engagement varies with the quality of the candidate–job match. Disclosure can change a jobseeker's expectations about an opportunity, and those expectations may differ with how well the jobseeker is matched to it. AI disclosure could therefore change the composition of the engaged pool, not just its size. The direction of that selection is not obvious in advance, and the average engagement effect alone cannot show which candidates are more or less likely to engage.

There are several reasons to expect the disclosed source of a match to alter engagement and the pool of potential candidates that forms. Being told that a recommendation was produced by AI rather than by a person can change what jobseekers infer about the matching system, for example: (A) how precisely it recognizes candidate–job fit, (B) what information it uses, and (C) how broadly opportunities are distributed, and therefore how much competition an invitation implies. Jobseekers may also (D) hold different attitudes toward being matched by AI than by a person. Section 2 develops these four channels. These beliefs and attitudes can change willingness to engage, and may do so differently for better- and worse-matched candidates. What matters behaviorally, then, is not "AI" in the abstract, but what its use implies about the matching system. If jobseekers respond to those features, the system does not simply operate on a fixed pool; it can help determine who enters in the first place. Platforms therefore choose not only how matching is performed, but what participants are told about it—a choice increasingly shaped by transparency requirements in algorithmic hiring (New York City Council, 2021; European Parliament and Council of the European Union, 2024).

We study this with a field experiment on a hiring platform at a large R1 university that matches its students to high-skilled, short-term project positions. The 4,465 student jobseekers received 4,534 individualized emails, each recommending a specific position. We randomly varied whether the email attributed the recommendation to an AI matching algorithm, a human employer, or no source, while holding the recommendation itself fixed. Independently, we randomized the wage. Because the matching ran before assignment and only the disclosed source varied, behavioral differences across the three arms identify responses to the attribution rather than to differences in the underlying match.

The research design required solving two central problems. First, we had to observe the full risk

set: not only who engaged, but which plausible candidates could have engaged and did not. Hiring data ordinarily begin with the people who apply. We therefore assembled the candidate pool and measured its attributes before any recommendation was sent. Second, both source attributions had to be truthful. The procedure involved hiring managers and platform staff specifying job-relevant keywords, and an algorithm searching for those terms in jobseekers' public LinkedIn profiles. The same recommendation therefore combined human judgment with algorithmic execution and could truthfully be attributed to either source. We measure engagement as whether a candidate clicks through to inspect the recommended position, the discovery step preceding any application. A companion lab experiment supplements the field design for probing mechanisms, measuring how stated beliefs and attitudes differ under the same disclosure wording.

In the field experiment, AI disclosure moved two margins at once. Engagement fell by 2.3 percentage points, about a quarter of the 8.9 percent control-group rate ($p = .017$), and human matching was statistically indistinguishable from no disclosure. But the loss was not evenly distributed. Among the best-matched quarter of candidates, engagement fell from 13.1 percent under human matching to 6.5 percent under AI, a decline of about half, while among the bottom quarter it did not fall at all. All four quartiles consist of candidates who already met the employer's stated criteria. Under human matching, engagement rises with match quality; under AI disclosure, that relationship largely disappears. Holding candidates and recommendations fixed, AI disclosure therefore lowers engagement with a given posting and, relative to human matching, leaves the engaged pool worse matched. The aggregate decline does not reveal which candidates changed behavior: here it is concentrated among the better matched. The decline itself is measured per posting: the same jobseekers may simply be spreading thinner effort across more of them, in which case total applications could rise even as engagement with each falls. Our design cannot separate those readings; Section 6 takes up what the distinction implies.

We explore four possible mechanisms using the two experiments together. The evidence aligns most closely with (A) expectations of lower match quality under AI, is weaker for (B) the information the recommendation source is expected to read and for (D) attitudes toward it, and is absent for (C) competition, though it does not identify a unique mechanism. Better-matched candidates respond the most to AI disclosure, a separate comparison of master's and undergraduate computer-science candidates for the same posting points in the same direction, and in the lab AI disclosure lowers perceived match quality and hiring probability. Jobseekers also expect AI to rely less on referrals, networks and soft skills, and report lower trust and comfort with AI matching, although the field cannot separate these responses from the match-quality account. We find no evidence that disclosure increases perceived competition. However, limitations of our design also mean we cannot rule out a competition response.

The broader implication is that matching technologies do not merely rank a fixed pool more or less accurately. By changing what participants know, what search costs, and how they perceive the competition, they can change who enters that pool in the first place. A platform therefore

cannot judge a matching system by the accuracy of its recommendations alone: the pool it works with is partly a product of the system itself. The particular direction we observe is contingent on how jobseekers interpret the technology and matching process in this setting. More generally, participants respond to features they infer about how matching works and how they are likely to fare under it. As technologies, implementations and attitudes evolve, those inferences—and therefore the direction of the response—may change, while the platform-design problem remains. If scalable matching changes both the number of opportunities encountered and who engages with each one, improvements in prediction need not translate mechanically into improvements in realized matching.

The paper contributes first to research on the design of matching platforms, by showing that the matching architecture can affect the composition of the population on which matching subsequently occurs. We study that problem in a labor market, where prior field experiments show that information supplied by the platform changes jobseeker behavior, though not that it reallocates search in the same way: Gee (2019) finds higher application rates without substantial redirection, whereas Fradkin et al. (2026) find meaningful reallocation across opportunities (on the employer side, see Horton, 2017). Gee and Fradkin et al. vary what jobseekers are told about demand for a job, whereas we vary what they are told about how the match itself was produced, and show that this changes not only engagement but who engages. This connects to directed-search models, in which workers sort across opportunities given their perceived chances of success (Moen, 1997; Shimer, 2005): our results suggest that the matching system itself can alter those perceptions and hence the pool that enters. Finally, work on algorithmic decision-making shows that disclosing AI can change behavior and stated attitudes (Keppeler, 2024; Tong et al., 2021; Avery et al., 2023; Xiong and Kim, 2025), and recent field evidence shows that a truthful disclosure can move labor-market behavior even when the underlying recommendation or ranking is held fixed (Filippas et al., 2025). Keppeler (2024), in particular, shows that AI disclosure can reduce engagement with job recommendations. What that work leaves open, and what we show, is that disclosure also changes *who* engages: it changes how participation varies with candidate–job match quality, and so the match-quality composition of the engaged pool.

The remainder of this paper is structured as follows. Section 2 sets out several reasons why disclosing AI rather than human matching could change jobseekers' willingness to engage, and why that change could differ with the quality of the match. Section 3 describes the field experiment's design and data. Section 4 presents the two main results: fewer jobseekers engage, and the decline is concentrated among the better matched. Section 5 explores four possible mechanisms using evidence from both the field and companion lab experiments, finding evidence consistent with three of them and narrowing the plausible interpretations without identifying the mechanism. Section 6 discusses implications and limitations.

2. AI MATCHING AND JOBSEEKER ENGAGEMENT

The comparison throughout is between a match attributed to an algorithm and the same match attributed to a person: what differs is the jobseeker's belief about who or what produced it, not the match itself. AI disclosure could plausibly affect two distinct dimensions of engagement: its *level* and its *composition*. The first concerns how many jobseekers engage; the second concerns how engagement is distributed over candidate–job match quality — whether relatively better- or worse-matched jobseekers engage.

A simple engagement condition helps organize these possibilities. A jobseeker engages with a recommended posting when its expected payoff exceeds the value of the next-best use of the same attention and effort:

$$p \cdot W - c \;\geq\; \delta,$$

where p is the jobseeker's subjective probability of being hired conditional on engaging, W is the value of the position, c is the direct cost of inspecting and pursuing it, and δ, the jobseeker's engagement threshold, is the value of devoting the same time and attention elsewhere. Attention is genuinely scarce: active jobseekers spread it across the many opportunities they might pursue, devoting on the order of half an hour a day to search (Krueger and Mueller, 2010), so engaging any particular posting is a costly and selective act. AI disclosure can affect this decision through four distinct channels, which we consider in turn.

These channels are analytically distinct but need not operate independently in practice: a particular matching technology or platform architecture may move several of them at once. [1]

A. Match quality. AI disclosure may change jobseekers' expectations about how accurately their match quality will be recognized, an expectation we represent as entering through the subjective probability of being hired, p. If jobseekers expect AI to be a less precise evaluator than a person, subjective hiring probabilities become less strongly related to match quality: better-matched candidates become less confident that their match quality will be recognized and expect a lower p, while worse-matched candidates may become less certain that poor match quality will count against them and expect a higher p. Because a jobseeker engages only when $p \cdot W - c \geq \delta$, a lower p can push better-matched candidates below the threshold, while a higher p can pull worse-matched candidates above it. Engagement therefore becomes less strongly related to match quality. (The prediction reverses if jobseekers instead expect AI to be the *more* precise evaluator.) We focus on expected precision as the simplest case. Other expectations about how algorithmic evaluation maps candidate characteristics into hiring probabilities—for example, that it systematically favors or disfavors particular candidate types—could generate different patterns of selection.

The predicted fall in engagement among better-matched candidates draws on the statistical-

[1] The mapping from "AI" to these features is itself evolving and depends on both technological capability and where automation enters the matching process (Autor, 2014; Tambe et al., 2019; Bommasani et al., 2021; Le Barbanchon et al., 2023).

discrimination tradition (Phelps, 1972; Aigner and Cain, 1977; Lundberg and Startz, 1983; Frankel and Kartik, 2019): an evaluator expected to be less informative compresses subjective hiring probabilities toward the middle, much as a Blackwell garbling reduces the information carried by a signal (Blackwell, 1953). That compression lowers the relative engagement of the better matched when three conditions hold: candidates have some sense of their own match quality, the employer's hiring decision is increasing in match quality, and enough candidates sit close to their engagement threshold for a change in p to move them across it.

B. Codified vs. soft and tacit inputs. AI disclosure may also change jobseekers' expectations about the information feeding a match prediction. An algorithm conditions its prediction on information available to the model, whereas a human evaluator may also observe information outside that data set, including soft or tacit information conveyed through referrals, reputation, endorsements, or informal assessment (Granovetter, 1995; Burks et al., 2015). Jobseekers may therefore expect an algorithm to rely relatively more on recorded and codified information and less on soft or tacit information (Tambe et al., 2019).

If jobseekers expect AI-generated matches to draw less on the soft sources, candidates who expect those sources to favor them should expect less of the evidence in their favor to be read, which we represent as a lower probability of being hired, p, when the match is attributed to AI than when it is attributed to a person, and they should engage less as a result. Candidates whose qualifications are carried by the codified sources face the opposite movement. The disclosure does not touch the value of their other, unlabelled options, so a recommendation source expected to weigh codified information more heavily raises p for them on this posting relative to those alternatives, and they should engage more. The channel therefore pushes in both directions on the level.

Whether this mechanism disproportionately affects better- or worse-matched candidates depends on whether their advantages are carried more by soft or tacit information than by codified information. It therefore predicts a composition effect when reliance on these information sources varies with match quality, but does not predict its sign.

C. Scale of competition. AI disclosure may also change how many competitors a jobseeker expects to face. Because jobseekers do not observe the full field invited to a posting, they form that expectation partly from how the matching process is described, and an algorithm and a person may suggest fields of different size. Automated matching may plausibly read as broader outreach, whereas more targeted matching may suggest a smaller candidate pool.

That inference matters because hiring is a tournament: candidates compete for a limited number of positions, so conditional on match quality, a jobseeker's chance of being hired depends partly on the expected field of competitors (Lazear and Rosen, 1981; Gee, 2019; Fradkin et al., 2026). We represent a wider field as a lower p; it could as readily raise what it takes to stand out, which would act on c. A larger expected field therefore lowers the subjective probability of being hired, p, and with it engagement; a smaller expected field raises both.

Whether this mechanism disproportionately affects better- or worse-matched candidates depends on how additional expected competition changes incentives across match levels. Tournament theory and evidence show that responses to additional competitors can vary with contestant ability and need not be monotonic, discouraging some contestants while strengthening incentives for others (Moldovanu and Sela, 2001; Boudreau et al., 2016; List et al., 2020). That work concerns effort conditional on entering a contest, whereas engagement is the entry decision itself; the same heterogeneity can arise at that margin when entering is costly (Fullerton and McAfee, 1999). The mechanism therefore predicts a composition effect when the response to expected competition varies with match quality, but does not predict its sign. Throughout, the mechanism concerns the field jobseekers *expect* to face, which need not be the field a platform actually invites.

D. Attitudinal response. AI disclosure may also affect engagement through an attitude or preference toward the recommendation source itself, independent of any belief about the likely outcome of engaging. A jobseeker may care about whether a machine or a person produced the recommendation even when expected outcomes are held fixed. This is a direct preference rather than an expectation about p, and we represent it as a shift in the cost of engaging, c. The human–AI interaction literature records both algorithm aversion and algorithm appreciation, with aversion the more common, though mostly outside labor markets and at lower stakes than a job search (Dietvorst et al., 2015; Logg et al., 2019; Burton et al., 2020; Diederich et al., 2022; Jussupow et al., 2023; Turel and Kalhan, 2023; Qin et al., 2025).

A shared aversion would lower the *level* of engagement and a shared appreciation raise it. Whether this mechanism disproportionately affects better- or worse-matched candidates depends on how attitudes toward AI vary with match quality, or on where candidates at different match levels sit relative to their engagement thresholds. It therefore predicts a composition effect under either condition, but does not predict its sign.

Three of the four channels enter through p, a jobseeker's own odds of being hired: two concern how AI shapes the matching process itself, and one the tournament that forms around the position. The fourth enters through c, the direct cost of engaging. A jobseeker who revises p on any of these grounds and engages less is responding to expected return, and may be doing so with fully rational, even Bayesian, expectations; the response requires no bias. The predictions below require only that jobseekers act on those expectations, not that they be correct; Section 6 returns to whether they may nevertheless be warranted or borne out in equilibrium.

2.1 Predictions for Engagement and Selection

The preceding discussion gives several reasons why disclosing AI matching could affect both how many jobseekers engage and which jobseekers do so. We now consider what those channels imply when taken together.

The arguments above do not imply a fixed effect of AI disclosure. Its direction depends on what the technology does, what is disclosed about it, and the attitudes jobseekers bring to both; as those

change, so may the response. An AI label matters because it can change what jobseekers infer about the matching system and how they regard the recommendation source, not because "AI" carries a fixed behavioral meaning of its own.

This contingency does not make behavior arbitrary within a setting. Conditional on how jobseekers read the matching system, the channels above apply systematically: expectations about matching precision, information inputs and the field of competitors enter through the expected return to engaging, and attitudes toward the recommendation source enter through its cost. Where those responses vary with match quality, they change which jobseekers engage as well as how many.

To derive directional predictions for this study we therefore adopt a benchmark drawn from the empirical setting and the state of AI matching at the time. Under it, jobseekers are taken to interpret AI matching as lower-cost and more scalable, and therefore as reaching a broader candidate pool; as less precise than human judgment in recognizing match quality; as more reliant on codified information; and as carrying some degree of aversion. Table 1 works through what the channels above imply under that benchmark.

Table 1: Channels Through Which Disclosure Alone Can Change Engagement and Its Composition, and Their Direction Under a Benchmark

Mechanism	*What it moves*	*Effect on level*	*Effect on composition*
Match quality	Expected precision of match evaluation (p)	Ambiguous	Flattens: engagement falls among the best matched and rises among the worst matched
Codified vs. soft and tacit inputs	Which information the evaluation is expected to draw on (p)	Ambiguous	Ambiguous: depends on whether reliance on soft and tacit information rises or falls with match quality
Scale of competition	Expected set of rivals, hence the probability of being hired (p)	Lowers	Ambiguous: depends on how the response to added rivals varies with match quality
Attitudinal response	Non-pecuniary cost of engaging (c)	Lowers	Ambiguous: depends on how attitudes vary with match quality, and on where candidates sit relative to their engagement thresholds

Note. Entries compare a match attributed to AI with the same match attributed to a person under the benchmark assumptions in the text. The channels locate each mechanism within the engagement condition; they do not explain how expectations or attitudes form. *Flattens* means that engagement becomes less strongly related to match quality; *ambiguous* means that the mechanism does not sign the net effect. Composition entries refer to how disclosure effects vary with candidate–job match quality.

The channels differ in how directly they imply a composition effect. If AI is expected to be the less precise evaluator, the match-quality channel predicts a flatter relationship between match quality and engagement. The other three channels change composition only if their effects vary systematically with match quality — for example, if better- and worse-matched candidates differ in

whether their advantages are carried by soft versus codified information, in how they respond to added competition, in their attitudes toward AI, or in where they sit relative to the engagement threshold. If we found no composition effect, that would count as evidence against the match-quality channel, provided enough candidates sit near their engagement threshold for the channel to show up in behavior. It would count against the other three channels only if the relevant source of heterogeneity were also present.

Under this benchmark, we expect AI disclosure to reduce overall engagement. A wider expected field and aversion work directly in that direction. Lower expected precision and a shift in which information is read both cut in opposite directions across the match-quality distribution rather than delivering an unambiguously negative aggregate effect. Taken together, we treat lower engagement as the natural directional prediction under this benchmark.

The prediction for composition is less clear. Lower expected precision implies relatively greater withdrawal among the better matched, but the effects of information inputs, competition, and attitudes can vary across match levels in either direction. We therefore do not sign the net composition effect ex ante, asking instead whether AI disclosure changes the relationship between match quality and engagement.

3. THE FIELD EXPERIMENT: SETTING AND DESIGN

Studying how disclosure affects both the level and composition of jobseeker engagement requires more than randomizing a label. The setting must make the relevant candidate pool observable before anyone acts, record an upstream engagement decision before application or employer screening, vary source attribution while holding the underlying recommendation fixed, ensure that alternative source descriptions are truthful, and measure candidate–job match quality before treatment for both those who engage and those who do not. Exploring why any effects arise presents an additional challenge, because the mechanisms developed in Section 2 operate largely through latent beliefs and expectations: about match quality, the information used by the recommendation source, the field of competitors, and the source itself. Field behavior can reveal patterns consistent with these channels, especially through heterogeneity in observable candidate characteristics, but actions alone generally do not identify the expectations that produced them. We therefore use the field experiment[2] to identify effects on engagement and selection and to probe observable implications of the proposed mechanisms, while Section 5 examines the mechanisms using the field data jointly with a companion lab experiment that measures directly how the same disclosure changes stated beliefs and attitudes. Among the empirical requirements, two were especially difficult: making AI and human source attributions both randomized and truthful, and observing the full risk set, including candidates who never engaged. This section elaborates the setting and design that make these comparisons possible.

[2]Both experiments were approved by the relevant institutional review boards. Details will be provided after the blind-review process.

3.1 The Matching Process and the Risk Set

We partnered with a university-based platform for high-skilled, short-term, and part-time project roles. The postings are real paid positions, and candidates vary in how well they match the one they were sent. We used the eight postings the employer advertised over a concentrated three-week period, chosen because they ran contemporaneously as three-week advertisements: they spanned cryptocurrency applications, full-stack engineering, data architecture, marketing strategy, video production, and psychology projects. Embedding the experiment in a live platform required one modification to the platform's usual process: recommendations were produced by a simpler matching rule than the platform normally uses. Hiring managers and platform staff jointly defined job-specific keywords capturing the relevant fields of study, technical skills, and prior experience, and a Boolean algorithm matched those keywords against the public LinkedIn profiles of all the university's students.

The coarseness was deliberate, for two reasons. The first is that it kept both labels truthful. Because the same recommendation was genuinely produced by a human step and a machine step, we could describe it as either AI- or human-generated and randomize only which source was disclosed. The second is that it produced a large pool in which every candidate was relevant to the posting they received, while still varying considerably in how well they answered it, and it made that pool observable in full before any email went out.

Eligibility was based on coarse relevance criteria rather than strength of match. The sequence was as follows. Working with the institution, we began from its internal records of the students to whom the opportunity was open and who also had a public LinkedIn profile. A jobseeker was then eligible for a posting if their field of study, technical skills or prior experience made them broadly relevant to it, as determined by the keyword rule above: anyone with matching education or expertise qualified. Eligibility was binary: every candidate who cleared the relevance bar was included in the risk set and received the recommendation, with no further ranking or discretionary selection within that set. The keyword rule also determined which posting each jobseeker was matched to, so no discretionary assignment could have been correlated with how well a profile answered a posting. The resulting risk set comprised 4,465 jobseekers across eight postings.[3][4]

The pool that the platform regards as the relevant pool of jobseekers and the pool of jobseekers who are in fact engaged with the postings are not the same. Of the 4,341 eligible candidates with a match-quality score, 348 engaged, or 8.0 percent, by clicking through to learn more about the posting. On average, over the entire sample, those who engaged were better matched than the overall pool, as can be particularly seen in panel B of Figure 1, where the proportion engaged increases with match quality.

[3] For all but 68 jobseekers there was a single job recommendation; 67 received two and one received three, and each was assigned the same arm on all of theirs. Clustering the standard errors by jobseeker rather than by recommendation changes every standard error in Tables 5 and 6 by less than one percent, and no p-value materially.

[4] On the later match-quality measure, the assigned posting is the highest-scoring of the eight for 28.9% of candidates, compared with 12.5% by chance; its mean rank is 3.2, compared with 4.5 by chance. Thus the coarse relevance screen loads positively on relevance, as intended, without mechanically selecting the highest-scoring posting.

<Figure 1>

3.2 Randomization and Treatments

Disclosure was randomized jobseeker by jobseeker within each posting's risk set, once jobseekers had been sorted to the posting relevant to them, with equal allocation targeted across the three arms; the wage was randomized independently of it. Because the postings ran contemporaneously, assignment was carried out at once rather than as postings launched.

The disclosure treatment. The jobseekers were randomized into one of three conditions related to prediction source; Appendix C reproduces the complete recommendation email and its randomized text. The control group provides no information about the recommendation source. The AI matching treatment includes a line that says, "Recommendation source: A.I. matching algorithm using publicly available LinkedIn profiles." The human matching treatment includes a line that says, "Recommendation source: The employer generated a list of candidates by reviewing publicly available LinkedIn profiles." Both descriptions referred to genuine components of the hybrid process: employers specified the criteria used to identify relevant candidates, and the algorithm applied those criteria to the profiles.

To accommodate varied interpretations of machine-based decision-making, we deliberately included both "AI" and "algorithm" in the AI treatment label.[5] At the time of the experiment, rule-based systems, natural language processing, and machine learning all coexisted alongside emerging large language models. Including both terms was a way of embracing the range of meanings and expectations jobseekers might hold, and the framing is intended to generalize beyond the state of the technology at that moment.

Each email was sent from a university address. Each email included a job title, a brief one-line description, an expected wage for the jobseekers assigned to one of the seven wage levels described below and no wage line for those assigned to the no-wage control, and a link to the platform where further information would appear on the position and where it would be possible to apply. Because these roles were not advertised through any other channel, this recommendation email was effectively the only way candidates would learn about the opportunity and act on it, making the click-through both the sole discovery mechanism and the necessary first step toward applying.

These three arms support three distinct comparisons, and we report all of them throughout because they answer different questions rather than estimating one quantity with varying precision. *AI versus Human* holds fixed that a source is named and varies only what that source is, so it isolates

[5]Computer science researchers Narayanan and Kapoor (2024) describe this situation in the first pages of their book: "Imagine an alternate universe in which people don't have words for different forms of transportation—only the collective noun 'vehicle.' They use that word to refer to cars, buses, bikes, spacecraft, and all other ways of getting from place A to place B. Conversations in this world are confusing. There are furious debates about whether... vehicles are environmentally friendly, even though... one side of the debate is talking about bikes and the other side is talking about trucks. There is a breakthrough in rocketry, but the media focuses on how vehicles have gotten faster... Now replace the word 'vehicle' with 'artificial intelligence,' and we have a pretty good description of the world we live in."

the response to the disclosed source attribution; it is the contrast the paper's mechanisms are about. *AI versus the control group* compares a named algorithmic source against unspecified provenance, which is the choice a platform actually faces when deciding whether to disclose at all. Two things differ across those two arms at once: whether any source is named, and whether the source named is an algorithm. A difference between them could therefore come from either. The human arm helps distinguish these possibilities: engagement under human matching is statistically indistinguishable from the control group, so we detect no effect of naming a human source on overall engagement. *AI versus the pooled non-AI arms* aggregates the two, gaining power at the cost of specificity.

The wage treatment. In addition to the disclosure treatment, we implemented a second orthogonal wage treatment, in which the expected wage was randomly drawn from a range of $16 to $28 per hour, in $2 increments, while a control group received no wage information. Randomly assigning the wage was feasible given institutional norms, where job postings often listed an illustrative range but the final rate was determined based on a candidate's education level and program of study. Manipulating wage expectations alongside disclosure supplies a within-experiment comparison: an ordinary change in the value of the offer, against which the disclosure response can be read.

3.3 Data and Variables

Our dataset combines information from the platform's customer relationship management (CRM) system, LinkedIn profiles, and educational data matched to standardized field classifications. The data are recommendation-level: 4,534 observations on the 4,465 jobseekers who were keyword-matched to a posting and received a recommendation, divided evenly among the AI matching treatment, the human matching treatment, and the control group. Main variable descriptive statistics appear in Table 2. Table 4 reports correlations among the match-quality measures and the jobseeker-record attributes that might stand behind them. Section 5 introduces further variables specific to its own analyses.

<Table 2>

Our main dependent variable is a jobseeker's *engagement*: the decision to spend any effort or attention on considering a position. We capture it as the click-through from the recommendation email to the posting, where the details of the position and the means of applying appear, and define *engagement* as an indicator equal to one when the recipient made that click. On this platform the recommendation is the sole channel through which a jobseeker discovers the posting (Section 3.2), so the click is the necessary gateway into the application process, and a shift in *who* makes it changes who enters the pipeline upstream of any employer screening. Engagement so measured is the initiation of search effort; we observe neither applications nor hires. We use *engaged pool* for those who click through, and do not equate clicking with applying.

Clicking through is a distinct stage between seeing an opportunity and applying to it: on a large job board a posting appears in far more searches than it is opened, and is opened several times over

for every application it receives (Marinescu and Wolthoff, 2020). Clicks are also directed, shifting toward occupations whose prospects hold up (Hensvik et al., 2021). A click therefore records that a posting cleared the jobseeker's attention. Engagement of this kind occurred for 8.2% of recipients across all treatment conditions.

Our main explanatory variables are the experimental treatments, represented by indicator variables for AI matching and human matching. The control group received the same email with no recommendation-source line, so no source was named; it is the omitted category and is estimated as the baseline level in the models that follow. The eight postings enter as fixed effects in the specifications indicated in each table, so that those comparisons are made within a posting.

Two classifications of technical background appear in the paper and they are not the same. The divisional grouping defined in the note to Table 2 is the university's own; it describes the sample, is checked for balance, and enters the regressions as controls. Section 5.5, which splits the sample on technical background, instead classifies each jobseeker as STEM or non-STEM from the education records on their profile, matched to standardized program codes following U.S. Citizenship and Immigration Services (2022), and says so where it is used.

The covariates are graduate degree, inferred female, year of first enrollment, the divisional grouping above, and the randomized wage. A *seniority index* summarizing how substantial a candidate's record is, taken as the first principal component of positions listed, LinkedIn connections, summary length, total profile length and graduate status, so that one number stands for five correlated attributes, enters separately in Section 4.2 as a rival attribute.[6] The index orders depth given relevance, since every candidate in the risk set was already matched on field, skills or experience. Section 4.2 uses it to ask whether the selection is about the candidate–posting pairing or about the person.

Measuring the quality of a match. How well a candidate suits the posting they were sent does most of the work in what follows. Recall that the risk set establishes only broad relevance: each candidate cleared a criterion of field, skills or experience, and a criterion everyone met cannot separate them. A finer measure is needed, one specific to the pairing: how closely this candidate's record answers this posting.

Our preferred measure is BM25, a ranking function from information retrieval that scores how closely a document matches a query. Thirty years after it was introduced it remains the reference baseline against which newer retrieval models are evaluated (Thakur et al., 2021). Here the query is the employer's posting text and the document is the candidate's pre-treatment LinkedIn profile.[7]

[6]Each of the five inputs is standardized before the decomposition. Positions listed is the count of titles on the profile, up to ten; summary length and total profile length are word counts; missing connection counts take the sample median. Because the sign of a principal component is arbitrary, the component is oriented to rise with positions listed, and is then standardized.

[7]BM25 was introduced in its modern Okapi form by Robertson et al. (1995), building on inverse-document-frequency weighting from Spärck Jones (1972); Robertson and Zaragoza (2009) give the modern derivation and standard parameterization. It is the default ranking function in the open-source search engines Lucene, Elasticsearch

BM25 has three properties that are useful here. First, rare and discriminating terms receive more weight, so a match on "cryptocurrency" counts for more than one on "business" (Spärck Jones, 1972). Second, the term-frequency component is saturating, so additional repetitions of a matched term receive progressively less weight. Third, document length is normalized within the scoring function. The last property matters because a leading alternative explanation is that high match scores simply reflect how much a candidate wrote rather than what was written.

For posting q and candidate profile d the score is

$$\mathrm{BM25}(q,d) \;=\; \sum_{t \in q} \mathrm{idf}(t)\, \frac{f(t,d)\,(k_1+1)}{f(t,d) + k_1\left(1 - b + b\,\dfrac{|d|}{\overline{|d|}}\right)},$$

where $f(t,d)$ is the count of term t in the profile, $|d|$ its length, $\overline{|d|}$ the mean profile length, and $\mathrm{idf}(t) = \log\!\big(1 + (N - n_t + 0.5)/(n_t + 0.5)\big)$ with n_t the number of profiles containing t. We use the conventional parameter values $k_1 = 1.5$ and $b = 0.75$ (Robertson and Zaragoza, 2009).

We also build multiple alternative measures of match quality, to assess how sensitive the composition result is to alternative constructions. Table 3 summarizes our preferred measure and each of these alternatives.

<Table 3>

<Table 4>

Table 2 reports balance across the three arms. Match quality, with which the composition result of Section 4.2 is an interaction, is balanced at $p = .22$ on the AI-versus-human contrast, as is the randomized wage. Sample sizes differ across tables because covariates differ slightly in availability.

3.4 Estimation Model

We estimate the level effect and the composition effect with two linear models on the recommendation-level data. The baseline specification of each carries neither posting fixed effects nor covariates; both are added in the columns indicated in Tables 5 and 6, and the estimates are stable across them. For jobseeker i facing posting j, the level effect is

$$y_{ij} \;=\; \alpha \;+\; \beta_{\mathrm{AI}}\,\mathrm{AI}_i \;+\; \beta_{\mathrm{H}}\,\mathrm{Human}_i \;+\; \mathbf{x}_i'\boldsymbol{\gamma} \;+\; \delta_j \;+\; \varepsilon_{ij}, \tag{1}$$

where y_{ij} is engagement, the indicators are the randomly assigned disclosure arms with the no-disclosure control group as the omitted category, $\mathbf{x}_i$ collects the pre-treatment covariates whose balance Table 2 reports, and δ_j are posting fixed effects. Because the arm was randomized within

and Solr.

each posting's risk set, β_{AI} is causal and the remaining terms should not move it; that they do not is reported rather than assumed.

Estimates related to the composition effect essentially add an interaction between treatment and the quality of match. Equation (2) is estimated on the AI and human arms, with human matching the omitted category. For comparisons with the no-disclosure control we estimate the analogous two-arm specification, with the control group as the omitted category. (Details of the estimates are given in Section 4.)

$$y_{ij} \;=\; \alpha \;+\; \beta_{\text{AI}}\,\text{AI}_i \;+\; \lambda\,M_{ij} \;+\; \theta\,(\text{AI}_i \times M_{ij}) \;+\; \mathbf{x}_i'\boldsymbol{\gamma} \;+\; \delta_j \;+\; \varepsilon_{ij}, \tag{2}$$

where M_{ij} is match quality, standardized over the pooled sample, so λ is how much engagement rises with match quality under human matching, per standard deviation, and θ (the coefficient reported throughout Section 4.2) is how much that relationship changes when the recommendation is attributed to an algorithm. A negative θ means the better-matched pull back more than the worse-matched. Every table states which.

4. MAIN RESULTS: LEVEL AND COMPOSITION OF ENGAGEMENT

In this section, we use the field experimental data, and the two specifications of Section 3.4, to test the predictions of Section 2.1: whether AI disclosure changes the level of engagement and the match-quality composition of the pool of candidates who engage. Section 4.1 reports that fewer jobseekers engage under AI disclosure. Section 4.2 shows that the decline is concentrated among better-matched candidates across multiple measures of match quality, leaving the engaged pool less well matched than under human matching on our preferred measure.

4.1 AI Disclosure and the Level of Engagement

The level effect. We begin by testing whether the level of engagement differs across arms. Model estimates regressing our engagement measure on each of the experimental arms (equation 1) are reported in Table 5, with standard errors in parentheses. Estimation is by a linear probability model,[8] with the jobseeker–posting recommendation as the unit of observation. Model (1) begins with the two treatment indicators and a constant, taking the no-disclosure control group as the omitted category.[9] Engagement is 8.9% in the control group and 9.0% under human matching, so human matching is not distinguishable from the control group (0.1 percentage points, $p = .92$). It

[8] Estimating the same specifications by logit and probit leaves the estimates and their significance unchanged. In the fully specified model the average marginal effect of AI disclosure is −2.5 percentage points under logit and −2.4 under probit, against −2.4 by least squares, and human disclosure remains indistinguishable from zero under all three ($p > 0.92$). We report the linear model because its coefficients are directly interpretable in percentage points and carry over to the interaction in Section 4.2 without further transformation.

[9] Conditioning estimates on just those jobseekers who were positively recorded as opening the emails delivers similar sign and significance of results, with the estimated coefficient on the AI treatment at −0.061 (s.e. 0.023) on 1,854 candidates. The experiment is unconditional on opening, and we report it that way throughout.

falls to 6.6% under AI. Against the control group that is a reduction of 2.3 percentage points (s.e. 1.0 percentage points, $p = .017$), or 26% of the control-group rate, with a 95% confidence interval running from 5% to 47%. Against human matching it is 2.4 points, or 27% of the human-matching rate ($p = .013$). Figure 2 plots the three arm means.[10]

<Figure 2>

Conditioning on covariates. Random assignment implies the estimate should survive conditioning on other variables. A full set of job-posting fixed effects (model 2) leaves the coefficients substantively unchanged. Model (3) adds the individual covariates: graduate degree, inferred female, year of first enrollment, and indicators for business programs, humanities programs and field not on file, with technical programs the omitted category. The AI coefficient estimate is unchanged throughout.

The wage benchmark. Model (4) adds the randomized wage, centered on its mean among the postings that showed one, along with a dummy for cases in which no wage was included, and carries no other controls. This also serves to provide an indication of the strength of the main treatment effects relative to this wage variation. Model (5) carries the wage alongside the job-posting fixed effects and the individual covariates. The estimate on the wage in model (4) is positive, the equivalent of 2.0 percentage points per \$10 change in wage, but it is not significant ($p = .11$). That point estimate is roughly the same size as the AI disclosure effect of 2.3 points, albeit not significant. Showing no wage at all makes little difference either: against a posting showing the average wage of \$22.50, the no-wage estimate is 1.1 points ($p = .29$). Neither estimate survives the full set of controls, as in model (5): the wage estimate falls to 0.66 points per \$10 ($p = .61$) and the no-wage estimate to 0.7 points ($p = .51$), while the AI coefficient holds at -2.4 points ($p = .015$). The reduction in engagement under AI disclosure remains essentially unchanged and statistically significant; in model (5), its point estimate is more than three times that associated with a \$10 increase in the wage.

Randomization inference. Candidates facing the same posting may respond alike. Clustering the standard errors would account for that, but the correction averages over the eight postings, and there are too few of them for that average to be reliable. We therefore use randomization inference, which reproduces the original within-posting assignment and requires no asymptotics in the number of postings (Young, 2019). The resulting inference is essentially unchanged: for the AI-versus-control comparison the randomization-inference p-value is .021, against .017 from the heteroskedasticity-robust standard error. Across the 14 contrasts reported in Appendix Table B2, the two never disagree about significance at conventional levels, and no p-value differs by more than .021 (Appendix Table B2). Appendix B.2 describes the procedure.

<Table 5>

[10]The equality of the control group and the human arm is not straightforward to read, since candidates told nothing about the source may have supposed a human one, leaving the two conditions different in wording rather than in belief. The direction is clear in either case: engagement under AI sits below both of the others, so what disclosure moves is engagement under AI rather than engagement under human matching.

4.2 AI Disclosure and the Composition of Engagement

Having found lower engagement under AI disclosure, we now test for composition effects in terms of the quality of match of those engaging. Model estimates regressing our engagement measure on the AI treatment, our measure of the quality of match and their interaction (equation 2) are reported in Table 6, with standard errors in parentheses. Recall that risk sets are constructed from individuals with relevant backgrounds (Section 3.1), so we are examining the effect of variation within this group.

The composition effect. Model (1) estimates the interaction on the 2,924 recommendations in the AI and human arms, taking human matching as the comparison group and carrying no controls or posting fixed effects. Match quality is BM25, defined in Section 3.3 and standardized, so every coefficient reads per standard deviation.[11] The coefficient on the interaction between the AI treatment and the quality of match is -0.028 (heteroskedasticity-robust s.e. 0.010, $p = .006$). This implies that engagement barely rises with the quality of match under AI disclosure: the slope is 2.9 percentage points per standard deviation under human matching and 0.1 points under AI. Panel A of Figure 3 plots this specification together with binned means of the data, which show the same pattern without imposing a functional form: engagement under human matching climbs from 5.4% in the bottom quarter of match quality to 13.1% in the top, while under AI it does not climb at all, and the top-quarter contrast between the arms against the bottom quarter is -8.4 percentage points ($p = .003$). The gap between the arms widens monotonically across the four quarters, at $+1.8$, -1.2, -3.4 and -6.6 percentage points. Against the no-disclosure control the interaction is -0.020 (s.e. 0.011, $p = .067$), the same sign and about two-thirds the size, while human matching is indistinguishable from the control ($+0.009$, s.e. 0.012, $p = .46$); the composition effect is therefore clearest against explicit human matching. The engaged pool itself is worse matched: mean match quality among those who engage is $+0.31$ standard deviations under human matching, $+0.19$ in the control group and $+0.04$ under AI. This is a different estimand from the interaction and a less precise one, because AI flattens the gradient, which worsens the pool, while also lowering the engagement rate, which raises selectivity among those who remain. The difference against human matching is -0.27 standard deviations ($p = .055$); randomization inference gives $p = .048$, and a Kolmogorov–Smirnov test rejects equality of the two engaged distributions ($D = 0.18$, $p = .044$). Against the control group it is not distinguishable from zero (-0.15, $p = .30$). Panels C and D of Figure 3 show this directly, as distributions rather than rates: the match quality of the whole risk pool, and of the candidates who engaged under each disclosed source. Under human matching the engaged pool is selected out of the risk pool; under AI it is not distinguishable from it. That holds on both measures, and more sharply within posting: against the pool, human matching gives $D = 0.18$, $p < .001$ pooled and $D = 0.22$, $p < .001$ within posting, while AI gives $D = 0.04$, $p = .99$ and $D = 0.06$, $p = .91$.

[11] Conditioning on the 1,205 jobseekers positively recorded as opening the email delivers the same sign and significance here too, with the interaction at -0.070 (s.e. 0.023).

<Figure 3>

Fixed effects and controls. Models (2) and (3) add posting fixed effects and then the full control set. The interaction is −0.028 in both ($p = .006$ and $p = .007$). Because BM25 is standardized over the pooled sample, posting fixed effects absorb differences in the level of engagement across postings but not differences in score distributions; measuring match quality as a candidate's rank within their own posting instead gives −1.0 percentage points per ten percentile points ($p = .004$), shown in panel B of Figure 3, so a candidate compared with others facing the same posting and holding the same field of study shows the same pattern. The relationship is close to linear: a quadratic term adds nothing (0.006, s.e. 0.007), and the binned means of Figure 3, panel A, track the fitted line. Nor is the result driven by any single posting: excluding each of the eight postings in turn leaves the interaction negative, with the smallest-in-magnitude estimate at −0.024. Randomization inference, as described in Appendix B.2, yields the same result, at $p = .006$.

Rival sources of treatment heterogeneity. Is the interaction really about candidate–job match quality, or could it instead reflect other candidate attributes correlated with our match-quality measure? Because match quality is measured rather than randomized, characteristics such as profile length, the number of positions listed, and seniority could themselves moderate the response to AI. Simply controlling for these attributes does not address that possibility; we therefore allow the AI effect to vary with each of them.[12]

Columns (4)–(6) of Table 6 add profile length, the number of positions listed, and the seniority index of Section 3.3 respectively, together with each attribute's interaction with AI disclosure. The index is the bundled version of this test: one number standing for five correlated attributes, so column (6) asks whether the bundle as a whole, rather than any single attribute, accounts for the pattern. The AI × match-quality interaction remains negative in every specification, at −0.022, −0.026 and −0.020, against −0.028 in the baseline. None of the rival-attribute interactions is individually distinguishable from zero, and neither is any of the implied changes in the match-quality interaction: the largest, when the seniority index enters, is 0.008 with a bootstrap 95% interval running from −0.019 to +0.002. None of these dimensions, then, accounts for the match-quality pattern.[13][14]

Alternative measures. A further check re-estimates Table 6, column (1), replacing the preferred measure with each alternative construction of Section 3.3 (Table 3). Figure 4 plots the four, and

[12]This is the familiar non-separability problem for treatment-effect heterogeneity defined by observed subject attributes (Holland, 1986; Sen and Wasow, 2016); see Chernozhukov et al. (2025, §2.2.3) and Boudreau (2026) for its bearing on experiments. There is no design remedy.

[13]The rival interactions are nonetheless all negative, and the attributes themselves carry positive level coefficients, two of the three distinguishable from zero, which is the same signature match quality shows. One plausible reading is that they carry components of match quality the preferred measure does not, since a more senior or more fully documented record may itself indicate a better match. Nothing here settles that.

[14]We ran the same check over every subject covariate available: profile length, positions listed, summary length, LinkedIn connections, the seniority index, graduate degree, inferred female, an unclassifiable first name, year of first enrolment, STEM field, bachelor-level study, master's-level study, doctoral study and a premium LinkedIn account. Each enters standardized, as a control and interacted with the AI treatment, alongside posting fixed effects. The AI × match-quality interaction remains negative and significant in all 14.

all four point the same way, to a more negative AI effect among the best-matched jobseekers. The interaction is -0.030 ($p = .008$) for plain word overlap, -0.023 ($p = .036$) for rare-word overlap, -0.028 ($p = .010$) for stated-skills overlap and -0.017 ($p = .12$) for meaning-based overlap, against -0.028 for the preferred measure.

<Table 6>

<Figure 4>

5. EXPLORING MECHANISMS: EVIDENCE FROM THE FIELD AND A COMPANION LAB EXPERIMENT

Section 2 developed four possible channels through which AI disclosure could alter engagement and its composition. The analyses that follow are exploratory rather than mechanism-identifying. The field experiment identifies the treatment effects documented in Section 4, but observed behavior alone does not reveal which beliefs or attitudes produced them. We therefore use the field data to examine observable patterns that would be consistent with the proposed channels, and a companion lab experiment (Section 5.1) to ask more directly whether the corresponding beliefs and attitudes move under the same disclosure contrast.[15] These analyses can establish whether particular interpretations are empirically plausible in this setting, but they cannot determine which channel caused the field selection effect.

The evidence is most consistently aligned with lower perceived match quality under AI. The lab also shows substantial differences in expected information inputs and in attitudes toward the recommendation source, although the field does not provide corresponding behavioral evidence that separates these interpretations. Neither experiment detects a significant difference in perceived competition, though neither measure is well suited to detecting one (Section 5.4). The evidence therefore sets competition aside and leaves match quality, information inputs, and attitudes as plausible channels, with match quality the best supported.

5.1 Companion Lab Experiment

Before proceeding to the analysis, we first describe the companion lab experiment. It ran at a second university.[16] There, 178 students were recruited through the behavioral laboratory across eight scheduled sessions, as an optional add-on to a separate and unrelated study, for a flat \$5. They average 19.7 years of age and range from 18 to 26; 72% are women, 69% are in their first or second year, and 29% are in a STEM major, across 39 fields of study (this pool is therefore similar

[15] Throughout this section the comparison is AI against human matching. The lab has only those two conditions, and in the field the no-disclosure control and human matching behaved alike on average (Section 4.1), so this is the contrast the two experiments share. Field estimates against the control group are reported in footnotes.

[16] Its sample size was set by the eight scheduled sessions rather than by a power calculation. Because recall proved to be affected by the treatment itself, screening on it would condition on a post-treatment variable (Montgomery et al., 2018), so we report the full randomized sample as primary and the correct-recall subsample as a sensitivity check (Appendix Table B1).

to the field study's in being university students of much the same age and stage, but differs in being entirely undergraduate, where 17% of field jobseekers hold a graduate degree, and in being markedly more female, 72% against an inferred female share of 50%; Table 2).[17] Participants were randomly assigned between subjects, with 87 respondents assigned to the AI matching condition and 91 respondents assigned to the human matching condition.

The design involves subjects reading a vignette (Appendix A). Each subject is shown a recommendation email that a university is described as having sent to a fictitious student, "Alex," offering a part-time position, using the same email template as in the field, including the same recommendation-source wording. The subject is not the recipient: they read the email Alex received and then give their impressions of it across fourteen items, two of which ask what they would do in Alex's position. The only element randomized in this exercise is the line naming the source of the recommendation: an AI matching algorithm, or the employer reviewing profiles itself. There is no third condition: the lab omits the field's no-source control group, by design, because the field showed essentially no average difference between human matching and that control, and splitting a fixed sample two ways rather than three buys precision.

Respondents could not return to the vignette after advancing. They answered a source-recall check from memory on a page of its own.[18] The questions divide into the four mechanisms of Section 2, and Table A1 lists every one as administered, with its wording, its scale and its place in the order. The blocks ran in a fixed order for every respondent: what each source considered, then perceived fit, then hiring probability, then stated engagement. That is the order of Section 2's argument, and it was identical in both arms, so it cannot generate the differences reported below. The lab thus records what jobseekers make of the disclosure once they have considered it; the field experiment records what they do unprompted. Asked how likely they would be to click through, respondents fall from 4.80 under human matching to 3.82 under AI, a drop of 0.98 points on the seven-point scale (s.e. 0.24, $p < .001$), about a fifth of the human-matching mean. Asked how interested they would be in applying, they fall from 4.52 to 3.72, a drop of 0.80 points (s.e. 0.24, $p < .001$). The disclosure therefore lowers stated engagement in the lab as it lowers actual engagement in the field.

Figure 5 reports how the answers differ between the two arms, item by item and grouped by the mechanism each speaks to, and is referred to throughout what follows. Each mechanism is then taken up in turn, with the field evidence and the lab evidence on each.

5.2 Match Quality: Do Jobseekers Expect AI to Recognize It Less Accurately?

A first possible mechanism behind the level and composition effects documented in Section 4 is that jobseekers expect AI to recognize candidate–job match quality less accurately. This is because, in principle, the best matched have most to lose from an imprecise evaluator and therefore would have

[17] The two gender measures are not the same quantity: the lab asks, while the field infers from first names and classifies only 75.8% of candidates.

[18] 87.4% of AI-arm respondents recall the source correctly, against 57.1% in the human arm, so the AI label registered clearly and the human label did not.

a differentially larger AI treatment effect.

Field experiment evidence. As an exploratory proxy in the field, we compare candidates at different levels of training within a single posting, the computer-science position, the one posting for which the employer expressed a preference for master's-level candidates. Master's-level computer-science candidates tend to carry more training and more industry experience than undergraduates. They are also better matched on the paper's own measure: among computer-science candidates recommended that posting, master's candidates score 0.82 standard deviations higher on BM25 than undergraduates (s.e. 0.07, $p < .001$). Within that group, master's candidates engage at 18.5% under human matching but only 9.7% under AI, whereas undergraduates engage at roughly 7% regardless of condition (7.8% and 6.5%). Estimated within each group, the AI-versus-human effect is -8.8 percentage points among master's candidates (s.e. 4.1, $p = .033$) and -1.3 points among undergraduates (s.e. 2.0, $p = .52$); the difference between them is -7.5 points (s.e. 4.6, $p = .10$). The point estimates therefore line up with the match-quality account, although the between-group difference is imprecisely estimated.

Lab experiment evidence. The companion experiment directly measures self-reported beliefs, with each of several results related to match quality being consistent with lower expected match quality. Respondents were asked "*How good a fit do you think this recommendation is for Alex?*" on a seven-point scale. Despite the same candidate and same position, perceived match quality is lower under AI disclosure, by 0.34 points on the seven-point scale (s.e. 0.15, $p < .05$). Further, the perceived hiring probability was elicited with the question "*If Alex applies for this position, how likely do you think Alex is to get it?*" on a 0–100 slider. The AI treatment lowers this score by 7.9 points (s.e. 3.3, $p < .05$; Figure 5 and Appendix Table B1).

Taken together, the evidence aligns most closely with the match-quality channel, although neither experiment directly identifies lower perceived precision as the cause. This channel maps directly onto the composition pattern of Section 4: if AI is expected to recognize match quality less precisely, the best matched have most to lose, and engagement should rise less steeply with match quality under AI. The field comparison of master's and undergraduate candidates is suggestive of that pattern. The lab shows that AI disclosure lowers both perceived fit and perceived hiring probability.

5.3 Information Inputs: Do Jobseekers Expect AI to Read Different Information?

A second possible mechanism behind the patterns documented in Section 4 is that jobseekers expect AI to draw on different information inputs. For example, if some candidates hold relational capital that a human evaluator is more likely to draw on, they may be differentially put off by AI matching. Further, if those same candidates tend to be among the better matched, we would see the composition effects of Section 4. We test this explanation next.

Field experiment evidence. We identify candidates who should be differentially put off by AI matching by the size of the professional network they hold, measured as LinkedIn connections. A larger network means more people placed to refer or recommend a candidate, which is the

information Section 2 expects a human evaluator to use more than an algorithm would. The count only approximates this: a candidate may know many people and have few who would provide a referral. Connections are observed for every candidate, with a median of 185 and a mean of 293, and are balanced across arms (joint $p = .91$). Candidates with larger professional networks indeed tend to be among the better matched, on average (the correlation between network size and match quality is 0.22, $p < .001$; Table 4). Engagement rises with network size under both human matching (0.30 percentage points per 100 connections, $p = .16$) and AI (0.18, $p = .32$), though neither slope is distinguishable from zero. The negative differential under AI is not statistically distinguishable from zero either: the AI slope is 0.12 percentage points below the human slope ($p = .67$).[19]

Lab experiment evidence. The lab experiment is designed to assess directly what inputs jobseekers assume were used under each matching condition. It compares relational and codified inputs within respondent: "*To what extent do you think {the A.I. matching algorithm / the employer} considered each of the following when making this recommendation?*" Three relational inputs (referrals, networks and soft skills) are set against three codified ones (listed LinkedIn skills, coursework and past internships), presented in randomized order. Respondents believe AI weighs relational inputs far less than a human evaluator does. On the seven-point scale, referrals fall by 1.86 points, from 3.83 under human matching to 1.98 under AI (s.e. 0.24, $p < .001$); networks by 1.15 points, from 4.64 to 3.49 (s.e. 0.25, $p < .001$); and soft skills by 0.70 points, from 3.48 to 2.78 (s.e. 0.24, $p = .004$). The codified inputs behave differently. Neither coursework nor past internships differs between the two conditions: coursework moves by -0.08 points (s.e. 0.24, $p = .73$) and past internships by -0.01 points (s.e. 0.26, $p = .96$), each indistinguishable from no change at all. Listed LinkedIn skills move in the opposite direction from the relational items, rising by 0.56 points under AI (s.e. 0.22, $p = .012$). Averaged over the three items on each side, AI disclosure moves the relational inputs down by 1.39 points more than the codified ones (s.e. 0.193, $p < .01$).

Taken together, jobseekers may expect AI to read different information, although there is not strong evidence that this is driving the difference in engagement in the field experiment. The lab shows that respondents expect AI to rely substantially less on relational inputs. In the field, if this expectation produced the composition effect, engagement should fall more under AI among the well networked, who are also better matched on average. The field does not show such a pattern: the slope under AI is 0.12 points below the slope under human matching, a difference that is small and imprecisely estimated.

5.4 Scale of Competition: Do Jobseekers Expect a Wider Field of Rivals Under AI?

A third possible mechanism behind the patterns documented in Section 4 is that jobseekers expect there to be a wider field of rivals and greater competition under the AI disclosure treatment. Since hiring is a tournament, a larger expected field lowers everyone's odds of being hired, which is how this mechanism would produce the level effect of Section 4.1. To explain the composition effect, this

[19] In the no-disclosure control group, engagement rises by 0.49 percentage points per 100 connections ($p = .04$); the AI slope lies 0.31 points below it ($p = .30$).

mechanism would additionally require higher expected competition to discourage better-matched candidates more than worse-matched ones.

Field experiment evidence. We use the log number of candidates eligible for a posting as a posting-level proxy for potential competition, and report every slope in percentage points of engagement per log point; risk sets run from 59 to 1,589 candidates across the eight postings. In the control group, engagement is lower where the pool is larger (-2.55, s.e. 1.05, $p = .015$). Under human matching and under AI the slopes are both statistically indistinguishable from zero (-0.25, s.e. 1.04, $p = .81$, and $+0.33$, s.e. 0.90, $p = .71$) and from each other (difference $+0.58$, s.e. 1.38, $p = .67$). Both disclosure arms sit above the control-group slope, AI by $+2.89$ (s.e. 1.39, $p = .038$) and human by $+2.31$ (s.e. 1.48, $p = .12$). These slopes are hard to interpret for two reasons. The proxy varies only across the eight postings, so anything else that differs between postings and happens to be related to pool size would produce the same pattern. And it measures the number of candidates actually eligible, whereas the mechanism we discuss in Section 2 depends on the number a jobseeker expects.

Lab experiment evidence. The lab asks respondents directly how large a field they believe received the same recommendation: "*How many other candidates do you think received this recommendation?*", on a seven-point scale running from very few to very many. The perceived pool does not differ between the two conditions, standing at 5.80 under human matching and 5.83 under AI (difference $+0.03$, s.e. 0.19, $p = .90$). Responses in the human matching condition already sit toward the top of the range, so the AI matching respondents had little room to report more rivals even if they expected them.

Taken together, we therefore find no evidence for this mechanism in either experiment. Given the limits of both measures just noted, the result should not be read as ruling out smaller effects, or effects in other settings.

5.5 Attitudinal Response: Do Jobseekers Simply Dislike Being Matched by a Machine?

A fourth possible mechanism behind the patterns documented in Section 4 is that jobseekers simply do not want to be matched by a machine on the basis of attitudes per se, rather than on its effect on matching or the competition for the position. An attitudinal response could generate a composition effect either if attitudes toward AI vary with match quality, or if a common attitudinal shift moves candidates differently because they sit at different points relative to their engagement thresholds. The field data speak, imperfectly, only to the first.

Field experiment evidence. The field experiment records actions, not attitudes. The closest we can come is to compare STEM and non-STEM candidates, on the idea that technical training may shape how a person feels about being evaluated by AI. STEM status is the federal designation of the candidate's program, classified for 4,384 of the 4,534 recommendations, and balanced across arms (joint $p = .74$). Both groups engage less under AI than under human matching: the effect is -2.86 percentage points among STEM candidates (s.e. 1.29, $p = .026$) and -1.94 points among non-STEM

candidates (s.e. 1.53, $p = .20$). The two are not distinguishable from each other (difference -0.92 points, s.e. 1.99, $p = .64$), and they remain so when we control for the posting, for match quality, and for the proxies used in the preceding subsections.[20] The test is uninformative in both directions. It finds no difference, and a difference would have been hard to attribute to attitudes, since technical background covaries with much else. Attitudes toward the recommendation source are better measured directly, which is what the lab does.

Lab experiment evidence. The lab experiment measures attitudes toward the recommendation source with three items: *trust*, "*How much do you trust {the A.I. matching algorithm / the employer} to make a good recommendation?*"; *fairness*, "*How fair do you think this recommendation process is?*"; and *comfort*, "*How comfortable would you be with {an A.I. matching algorithm / an employer} making job recommendations for you personally?*", each on a seven-point scale. Relative to human matching, AI disclosure lowers trust by 0.81 points (s.e. 0.19, $p < .001$, from a human-matching mean of 4.13) and comfort by 1.41 points (s.e. 0.24, $p < .001$, from 4.73). Fairness moves least, and is the only one of the three not clearly distinguishable from zero (-0.32, s.e. 0.19, $p = .10$). With the candidate and the position identical across conditions, the lab shows a clear but uneven attitudinal response: what moves is trust in the recommendation source and comfort with it making job recommendations, and fairness less so. At the same time, trust is not cleanly separable from assessments of match quality: within arm, the correlation coefficient between trust and perceived match quality is 0.53. This correlation is one reason we do not treat the attitudinal and payoff interpretations as competing explanations that can be adjudicated.

Taken together, the field does not detect a difference between STEM and non-STEM candidates in the decline in engagement under AI. The lab shows lower trust in and comfort with the AI recommendation source, although these responses are not cleanly separable from beliefs about match quality.

<Figure 5>

6. DISCUSSION AND CONCLUSION

AI disclosure changes both the level and composition of engagement — and therefore the pool from which any later match to a job can be made. In our field experiment, attributing a recommendation to AI lowers engagement by about a quarter relative to both no disclosure and explicit human matching. More important, the decline is concentrated among the better-matched candidates. Among the best-matched quarter, engagement falls from 13.1% under human matching to 6.5% under AI, a decline of about half, while among the bottom quarter it does not fall at all. Relative to human matching, AI disclosure therefore changes not only how many jobseekers engage, but the match-quality composition of the pool available for subsequent matching.

[20] Compared against the no-disclosure control, the ordering reverses: the effect is -3.98 points among non-STEM candidates (s.e. 1.63, $p = .015$) and -1.39 points among STEM candidates (s.e. 1.24, $p = .27$), a difference of $+2.60$ points (s.e. 2.05, $p = .21$).

Using evidence from both the field and companion lab experiments, we explore the four possible mechanisms of Section 2 underlying the change in engagement and its composition. The evidence is most clearly consistent with (A) expectations of lower match quality under AI, which provides one plausible interpretation of the larger decline in engagement among the best-matched candidates. To a lesser extent, the evidence is also consistent with (B) jobseekers expecting AI to rely on different information inputs: respondents expect AI to place substantially less weight on relational information, while candidates with larger professional networks tend to be better matched in the field. The evidence also indicates (D) less favorable attitudes toward AI matching that may operate partly independently of expected payoffs. We find no detectable evidence for (C) the competition mechanism: although AI can make it inexpensive to identify and reach a much broader candidate pool, respondents did not infer a larger field of rivals under AI; the evidence does not rule out such a response in other settings.

AI disclosure can matter through what jobseekers infer about the matching system and how they regard the recommendation source. In the companion experiment, being told that AI produced a recommendation changes stated expectations about how matching is conducted—how precisely fit is recognized, what information is used, and how likely the candidate is to succeed—as well as attitudes toward the source. These responses can shape the incentive to engage. Narrowly, this qualifies the generality of the particular treatment effect we estimate: what AI disclosure implies, and therefore the direction and magnitude of the response, need not be stable across settings or over time. It will depend on the capabilities of the AI, the matching architecture and operating model in which it is embedded, the role of human judgment, and what and how the platform discloses. Closely related field evidence suggests these multiple factors play a role. Filippas et al. (2025) held a candidate's application and its position in the employer's list fixed and varied only what employers were told about it. Disclosing that the applicant had paid to promote the application raised the likelihood of being hired, whereas removing the platform's algorithmic "Best Match" label from every application did not significantly change employers' hiring or the response to the promoted applications. The disclosure that affected behavior conveyed a costly action by the applicant that employers could not otherwise observe. The broader and more durable implication is that, conditional on how participants understand these features of the matching system, their participation responses should be systematic and, at least in part, predictable. The general phenomenon is therefore not a fixed behavioral response to "AI," but participation shaped by what jobseekers infer about the matching system they are entering.

Relation to existing work. The field experiment mentioned above shows that what a platform truthfully discloses about an application can affect employer hiring even when the application and its position in the employer's list are held fixed (Filippas et al., 2025). Our contribution differs along three dimensions: the side of the market is the jobseeker rather than the employer, the stage is upstream entry rather than downstream evaluation, and we ask whether disclosure changes selection on candidate–job fit, not only average participation. A second recent experiment makes a closely

related point from the employer's side. Wiles and Horton (2025) offered employers AI-written first drafts of their job posts; posting volume rose while matches did not, largely because the additional posts came from employers with lower hiring intent, and partly because the posts themselves were less informative. An AI intervention there changes the composition of the opportunities entering a matching market. Our result is its counterpart on the other side: information about how the matching is done changes the composition of the jobseekers entering a particular matching process, and does so along candidate–job fit. Field experiments also show that platform-supplied information can alter job search, although its implications for how search is allocated differ across settings: Gee (2019) finds higher application rates without substantial redirection, whereas Fradkin et al. (2026) find meaningful reallocation across opportunities. Our result identifies another compositional margin: information about how the match was produced changes how participation varies with candidate–job fit. Search and matching models study how workers and vacancies meet given the frictions between them (Mortensen, 1986; Pissarides, 2000), and directed-search models let workers sort among opportunities given their expected returns (Moen, 1997; Shimer, 2005); our evidence shows that what workers understand about the matching architecture can alter those expectations before entry and thereby help determine the set of participants available for subsequent matching. Prior work shows that disclosing algorithmic involvement can alter behavior in settings ranging from job recommendations (Keppeler, 2024) to workplace feedback, where no entry decision is at stake (Tong et al., 2021). Broader work on algorithm aversion documents related responses across decision-making contexts outside labor markets (Burton et al., 2020). Our result shifts attention from the average response to its incidence: disclosure changes which candidate–job matches survive the participation decision.

Platform design, disclosure, and endogenous selection. Matching platforms are typically evaluated using the candidates who apply, are screened, or are hired. Yet the platform may already have shaped that pool by affecting who chose to engage. Candidates who do not engage leave no downstream record, so this selection is easy to miss. Our experiment observes that margin directly: holding the recommendation fixed, changing only its attributed source changes which candidates inspect and engage with the opportunity.

A platform is therefore designing more than a recommendation. A matching technology does not simply process a fixed pool more or less efficiently; the operating model in which it is embedded can also shape who enters that pool and how participation is distributed across potential matches. Disclosure is one way participants learn about that operating model: credible information about how candidates were identified, what information the recommendation source uses, how broadly an opportunity was distributed, or how much screening has already occurred can change what they expect from engaging. Once participation is understood to respond to such expectations, the natural design question is no longer simply how to improve the matching algorithm, but how the entire matching architecture shapes entry and selection. The underlying design problem is not specific to labor markets. Matching markets vary widely in operating model: dating platforms

differ in who can initiate contact and how broadly recommendations are distributed; school-choice and residency systems differ in how preferences are elicited and matches are centrally coordinated; labor platforms likewise vary in screening, outreach, and the role of human judgment. AI can be embedded in any of these architectures, and its effects on participation may depend as much on that operating model as on the algorithm itself. Those choices may affect not only how well a platform matches the participants it has, but which participants it has to match in the first place. The managerial implication is not that a platform should withhold what it does. It is that a recommender should not be optimized against a candidate pool assumed to be fixed: the quality of the recommendation, what the platform communicates about how it was produced, and the participation these induce are therefore choices that should be considered jointly. This also means that disclosure requirements need not be behaviorally neutral: information intended to make algorithmic systems more transparent can itself alter who enters the process, although our evidence does not determine whether the resulting selection is socially desirable.

Why better prediction need not produce better market matching. This endogenous response may also help explain why increasingly powerful matching technologies can coexist with rising application volumes without corresponding improvements in matching. Our experiment identifies engagement with a given candidate–job match, not how search is reallocated across the market. But if cheaper and more scalable matching allows employers to reach more candidates while jobseekers spread their attention across more opportunities, applications and search activity can rise even as attention to any particular match falls and becomes less concentrated among the candidates best suited to it. Later screening can select only among those who enter the pipeline; it cannot recover well-matched candidates who never engage. In increasingly high-volume online labor markets (Greenhouse Software, 2026; Monster, 2026), evaluating new matching technologies may therefore require evaluating the operating model and behavioral equilibrium they create, not simply the predictive accuracy of the algorithm holding participation fixed. Better prediction at one stage need not yield better matching if the resulting system also changes participation, attention, and the informational content of engagement.

Our evidence identifies an upstream participation response to how a matching system is understood, not downstream applications, hiring, or equilibrium market outcomes, and it does not establish how behavior would change when the matching technology itself changes. Future work can trace this response through to applications and hiring, and on to organizational and market outcomes, and can ask how different matching architectures—and the expectations they induce—shape participation and selection across settings.

The central implication is that a matching system does not merely rank the people who arrive at its door; it can help determine who arrives there. Evaluating matching systems therefore requires looking beyond predictive accuracy to the participation and selection their design induces. The pool on which a matching system operates is partly an outcome of the system itself.

REFERENCES

Aigner, D. J. and Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review*, 30(2):175–187.

Autor, D. H. (2014). Polanyi's paradox and the shape of employment growth. NBER Working Paper 20485, National Bureau of Economic Research, Cambridge, MA.

Avery, M., Leibbrandt, A., and Vecci, J. (2023). Does artificial intelligence help or hurt gender diversity? evidence from two field experiments on recruitment in tech. Monash University Discussion Paper No. 2023-09.

Blackwell, D. (1953). Equivalent comparisons of experiments. *Annals of Mathematical Statistics*, 24(2):265–272.

Bommasani, R., Hudson, D. A., et al. (2021). On the opportunities and risks of foundation models. Technical report, Center for Research on Foundation Models, Stanford University. arXiv:2108.07258.

Boudreau, K. J. (2026). Field experiments in the science of science: Lessons from peer review and the evaluation of new knowledge. In Veugelers, R. and MacGarvie, M., editors, *The Economics of Science*. University of Chicago Press. Pre-print circulated as SSRN Working Paper No. 6025114.

Boudreau, K. J., Lakhani, K. R., and Menietti, M. E. (2016). Performance responses to competition across skill levels in rank-order tournaments: Field evidence and implications for tournament design. *RAND Journal of Economics*, 47(1):140–165.

Burks, S. V., Cowgill, B., Hoffman, M., and Housman, M. (2015). The value of hiring through employee referrals. *Quarterly Journal of Economics*, 130(2):805–839.

Burton, J. W., Stein, M.-K., and Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2):220–239.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2025). Fisher–schultz lecture: Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. *Econometrica*, 93(4):1121–1164.

Diederich, S., Brendel, A. B., and Lichtenberg, S. (2022). Algorithm aversion in information systems research: A comprehensive literature review. *Business & Information Systems Engineering*, 64(1):47–64.

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1):114–126.

European Parliament and Council of the European Union (2024). Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence and amending regulations (ec) no 300/2008, (eu) no 167/2013, (eu) no 168/2013, (eu) 2018/858, (eu) 2018/1139 and (eu) 2019/2144 and directives 2014/90/eu, (eu) 2016/797 and (eu) 2020/1828 (artificial intelligence act). Official Journal of the European Union, OJ L, 2024/1689, 12 July 2024. ELI: http://data.europa.eu/eli/reg/2024/1689/oj.

Filippas, A., Horton, J. J., Parasurama, P., and Urraca, D. (2025). Preference signaling in labor markets: Evidence from a field experiment. Working paper.

Fradkin, A., Bhole, M., and Horton, J. J. (2026). Competition avoidance vs. herding in job search: Evidence from large-scale field experiments on an online job board. *Management Science*, 72(2):1305–1323.

Frankel, A. and Kartik, N. (2019). Muddled information. *Journal of Political Economy*, 127(4):1739–1776.

Fuller, J. B. and Raman, M. (2021). Hidden workers: Untapped talent. Technical report, Harvard Business School / Accenture.

Fullerton, R. L. and McAfee, R. P. (1999). Auctioning entry into tournaments. *Journal of Political Economy*, 107(3):573–605.

Gee, L. K. (2019). The more you know: Information effects on job application rates in a large field experiment. *Management Science*, 65(5):2077–2094.

Granovetter, M. (1995). *Getting a Job: A Study of Contacts and Careers*. University of Chicago Press, 2nd edition.

Greenhouse Software (2026). Hiring benchmarks 2026: Recruiting metrics and trends. Analysis of more than 6,000 companies and 640 million applications, 2022–2025.

Hansen, B. E. (2022). *Econometrics*. Princeton University Press, Princeton, NJ.

Hensvik, L., Le Barbanchon, T., and Rathelot, R. (2021). Job search during the COVID-19 crisis. *Journal of Public Economics*, 194:104349.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.

Horton, J. J. (2017). The effects of algorithmic labor market recommendations: Evidence from a field experiment. *Journal of Labor Economics*, 35(2):345–385.

Jussupow, E., Spohrer, K., Heinzl, A., and Gawlitza, J. (2023). A review of algorithm aversion in human–AI interaction. *Information Systems Frontiers*, 25(3):755–776.

Keppeler, F. (2024). No thanks, dear AI! understanding the effects of disclosure and deployment of artificial intelligence in public sector recruitment. *Journal of Public Administration Research and Theory*, 34(1):39–52.

Krueger, A. B. and Mueller, A. (2010). Job search and unemployment insurance: New evidence from time use data. *Journal of Public Economics*, 94(3–4):298–307.

Lazear, E. P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89(5):841–864.

Le Barbanchon, T., Hensvik, L., and Rathelot, R. (2023). How can AI improve search and matching? evidence from 59 million personalized job recommendations. SSRN working paper 4604814.

List, J. A., van Soest, D., Stoop, J., and Zhou, H. (2020). On the role of group size in tournaments: Theory and evidence from laboratory and field experiments. *Management Science*, 66(10):4359–4377.

Logg, J. M., Minson, J. A., and Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151:90–103.

Lundberg, S. J. and Startz, R. (1983). Private discrimination and social intervention in competitive labor markets. *American Economic Review*, 73(3):340–347.

Marinescu, I. and Wolthoff, R. (2020). Opening the black box of the matching function: The power of words. *Journal of Labor Economics*, 38(2):535–568.

Michael, J. (2007). 40000 Namen, Anredebestimmung anhand des Vornamens. *c't Magazin für Computertechnik*, 17:182–183.

Moen, E. R. (1997). Competitive search equilibrium. *Journal of Political Economy*, 105(2):385–411.

Moldovanu, B. and Sela, A. (2001). The optimal allocation of prizes in contests. *American Economic Review*, 91(3):542–558.

Monster (2026). Job application behavior report. Survey of 1,006 U.S. jobseekers fielded March 2026.

Montgomery, J. M., Nyhan, B., and Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62(3):760–775.

Mortensen, D. T. (1986). Job search and labor market analysis. In Ashenfelter, O. and Layard, R., editors, *Handbook of Labor Economics*, volume 2. North-Holland.

Narayanan, A. and Kapoor, S. (2024). *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*. Princeton University Press.

New York City Council (2021). Local law no. 144 of 2021: A local law to amend the administrative code of the city of new york, in relation to automated employment decision tools. Int. No. 1894-A; codified at N.Y.C. Admin. Code §§20-870 to 20-874. Enforcement by the NYC Department of Consumer and Worker Protection commenced 5 July 2023.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review*, 62(4):659–661.

Pissarides, C. A. (2000). *Equilibrium Unemployment Theory*. MIT Press, 2nd edition.

Qin, X., Zhou, X., Chen, C., Wu, D., Zhou, H., Dong, X., Cao, L., and Lu, J. G. (2025). AI aversion or appreciation? a capability–personalization framework and a meta-analytic review. *Psychological Bulletin*, 151(5):580–599.

Robertson, S. and Zaragoza, H. (2009). The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4):333–389.

Robertson, S. E., Walker, S., Jones, S., Hancock-Beaulieu, M. M., and Gatford, M. (1995). Okapi at TREC-3. In *Proceedings of the Third Text REtrieval Conference (TREC-3)*, pages 109–126. National Institute of Standards and Technology.

Sen, M. and Wasow, O. (2016). Race as a bundle of sticks: Designs that estimate effects of seemingly immutable characteristics. *Annual Review of Political Science*, 19:499–522.

Shimer, R. (2005). The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy*, 113(5):996–1025.

Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1):11–21.

Tambe, P., Cappelli, P., and Yakubovich, V. (2019). Artificial intelligence in human resource management: Challenges and a path forward. *California Management Review*, 61(4):15–42.

Thakur, N., Reimers, N., Rücklé, A., Srivastava, A., and Gurevych, I. (2021). BEIR: A heterogenous benchmark for zero-shot evaluation of information retrieval models. In *Proceedings of the 35th Conference on Neural Information Processing Systems, Datasets and Benchmarks Track*.

Tong, S., Jia, N., Luo, X., and Fang, Z. (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal*, 42(9):1600–1631.

Turel, O. and Kalhan, S. (2023). Prejudiced against the machine? implicit associations and the transience of algorithm aversion. *MIS Quarterly*, 47(4):1369–1394.

U.S. Citizenship and Immigration Services (2022). STEM designated degree program list. U.S. Department of Homeland Security.

Wiles, E. and Horton, J. J. (2025). Generative AI and labor market matching efficiency. Revise and resubmit, *Management Science*.

Xiong, Y. and Kim, J. K. (2025). Who wants to be hired by AI? how message frames and AI transparency impact individuals' attitudes and behaviors toward companies using AI in hiring. *Computers in Human Behavior: Artificial Humans*, 3:100120.

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, 134(2):557–598.

TABLES

Table 2: Summary Statistics and Balance Across Disclosure Arms

	Full sample			*Mean by arm*			
	Mean	*Std. dev.*	*N*	*Control Group*	*AI Matching*	*Human Matching*	*AI vs. Human p-value*
Engagement *[clicked through]*	0.082	0.274	4,534				
AI matching	0.334	0.472	4,534				
Human matching	0.337	0.473	4,534				
Opened the email	0.409	0.492	4,534	0.399	0.404	0.424	0.27
				(0.013)	(0.013)	(0.013)	
Match quality *[BM25]* †	6.289	4.377	4,341	6.192	6.438	6.236	0.22
				(0.113)	(0.118)	(0.115)	
Graduate degree	0.171	0.377	4,534	0.156	0.185	0.173	0.36
				(0.009)	(0.010)	(0.010)	
Inferred female‡	0.499	0.500	3,437	0.523	0.479	0.493	0.50
				(0.015)	(0.015)	(0.015)	
Year of first enrollment	2019.5	1.1	4,517	2019.5	2019.5	2019.5	0.62
				(0.02)	(0.03)	(0.02)	
Technical programs§	0.581	0.493	4,384	0.599	0.560	0.583	0.21
				(0.013)	(0.013)	(0.013)	
Business programs	0.242	0.428	4,384	0.235	0.252	0.238	0.35
				(0.011)	(0.011)	(0.011)	
Humanities programs	0.178	0.382	4,384	0.166	0.188	0.179	0.56
				(0.010)	(0.010)	(0.010)	
Wage *[$100s; zero if none shown]*	0.174	0.099	4,534	0.175	0.175	0.174	0.80
				(0.003)	(0.003)	(0.003)	
No wage shown	0.225	0.418	4,534	0.221	0.224	0.230	0.74
				(0.011)	(0.011)	(0.011)	

Notes. Cells are means; standard deviations in the full-sample columns, standard errors of the mean in parentheses under the arm columns. $N = 4{,}534$ (1,490 control, 1,515 AI, 1,529 human), varying by row where a variable is not observed for every candidate. The tabulated test is the *AI vs. Human* contrast, from a two-sided t-test of equal means with heteroskedasticity-robust standard errors. The covariates do not jointly predict assignment: regressing the AI indicator on all 10 of them gives $F = 0.91$ ($p = .53$) across the 3,044 candidates in the AI and human arms. No single-variable three-arm test falls below $p = .09$. The first three rows are the outcome and the arm indicators, so carry no balance figures; opening the email is not a pre-treatment covariate and does not enter the joint test. †Defined in Table 3; reported in raw units and standardized before estimation. §Programs follow the university's divisional structure: technical is computer sciences, engineering, science and health sciences; humanities programs include humanities, the social sciences, arts, media and design; business the business school. ‡Gender is inferred from first names using Michael (2007), which classifies 75.8% of candidates. The 150 candidates with no program code, and those whose name the dictionary cannot classify, carry their own indicators rather than being dropped.

Table 3: Alternative Measures of Candidate–Job Match Quality

Measure	Description
Match quality *[BM25]*	**The preferred measure.** It counts the employer's words that appear on the candidate's profile, but gives more weight to unusual words, progressively less weight to repeated mentions of the same word, and adjusts for how much the candidate wrote. We use the conventional parameter values $k_1 = 1.5$ and $b = 0.75$ (Robertson et al., 1995; Robertson and Zaragoza, 2009); The exact formula can be found in Section 3.3.
Plain word overlap *[word-boundary share]*	It simply counts the share of words in the job posting that appear anywhere on the candidate's profile. There is no weighting, length adjustment, or stemming. This makes it especially transparent, but also means that candidates can score higher simply because they wrote more—which is one possible alternative explanation for the composition result.
Stated-skills overlap *[curated word share]*	It uses the same word-overlap count as plain word overlap, but looks only at the parts of the profile in which candidates directly describe themselves: the headline, summary, listed skills, industry, certifications, and job titles. It excludes the free-text descriptions of previous jobs. Restricting the profile this way means 25% of candidates score zero, compared with 14% when the full profile is used.
Rare-word overlap *[TF–IDF cosine]*	It also measures shared words, but gives more weight to unusual words than to common ones. Unlike BM25, repeated uses of the same word continue to add weight. Technically, the score is the cosine similarity between the posting and profile after weighting words by inverse document frequency (Spärck Jones, 1972).
Meaning-based overlap *[semantic embedding]*	It compares meaning rather than exact wording, so two passages can score as similar even when they use different words. The posting is compared with each main part of the candidate's profile—the headline, summary, skills and certifications, and each job title and description—and the score is the maximum of those similarity scores. Similarity is measured with sentence-transformer embeddings using `sentence-transformers/msmarco-distilbert-base-v4`.
Match quality within posting *[BM25 percentile within posting]*	It uses the same BM25 score, but asks where a candidate stands among the candidates considered for the same posting rather than on an absolute scale. Because BM25 levels depend on a posting's own vocabulary and length, scores are not directly comparable across postings; this construction compares each candidate only with others who faced the same posting.

Table 4: Correlations Among Match-Quality Measures and Jobseeker-Record Attributes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Constructions of match quality														
(1) Match quality	1.00													
(2) Plain word overlap	0.73	1.00												
(3) Rare-word overlap	0.61	0.66	1.00											
(4) Stated-skills overlap	0.49	0.78	0.54	1.00										
(5) Semantic embedding	0.33	0.49	0.31	0.53	1.00									
Attributes potentially correlated with measures of match quality														
(6) Seniority index	0.43	0.51	0.09	0.42	0.38	1.00								
(7) Positions listed	0.34	0.36	0.08	0.27	0.40	0.69	1.00							
(8) Length of profile	0.40	0.51	0.08	0.38	0.31	0.87	0.54	1.00						
(9) Summary length	0.27	0.30	0.03	0.37	0.17	0.65	0.20	0.56	1.00					
(10) LinkedIn connections	0.22	0.30	0.09	0.24	0.23	0.65	0.39	0.39	0.18	1.00				
Other jobseeker characteristics														
(11) Graduate degree	0.16	0.17	0.02	0.11	0.13	0.46	0.08	0.29	0.19	0.25	1.00			
(12) Inferred female	0.04	0.05	0.07	0.05	0.02	-0.01	0.14	0.00	-0.04	-0.03	-0.17	1.00		
(13) Year of first enrollment	0.04	0.05	0.01	0.03	0.03	0.10	-0.13	0.05	0.07	0.07	0.41	-0.09	1.00	
(14) STEM field	-0.03	-0.20	-0.22	-0.25	-0.15	0.00	-0.14	0.02	0.04	-0.05	0.19	-0.15	0.14	1.00
Observations: 4,341 (all three arms)														

Notes. Pearson correlations, computed pairwise, so cell samples vary with the availability of each pair; the reported count is the maximum. Rows (1)–(5) are the match-quality measures of Table 3 that are built from text, in raw units; rank within posting is a re-referencing of BM25 and correlates with it by construction, so it is not repeated here. Row (6) is the seniority index of Section 3.3, the first principal component of rows (7)–(10) together with graduate degree (row 11); The three attributes in rows (6)–(8) enter Table 6, models (4)–(6), as rival interactions. Row (14) is the STEM classification of Section 5.5, built from program codes; it is not the technical-programs division of Table 2, which follows the university's own colleges.

Table 5: Effects of AI and Human Disclosure on Engagement

Dep. Var.:	*Engagement (click-through)*				
Model:	(1)	(2)	(3)	(4)	(5)
AI Matching	-0.023^{**}	-0.024^{**}	-0.024^{**}	-0.023^{**}	-0.024^{**}
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Human Matching	0.001	0.001	0.000	0.001	0.000
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Wage ($100s)				0.203	0.066
				(0.126)	(0.131)
No Wage Shown				0.011	0.007
				(0.010)	(0.010)
Constant	0.089^{***}	0.058^{***}	0.046^{***}	0.087^{***}	0.045^{***}
	(0.007)	(0.009)	(0.011)	(0.008)	(0.012)
AI vs. Human (reference: Human)	-0.024^{**}	-0.024^{**}	-0.024^{**}	-0.024^{**}	-0.024^{**}
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Job Posting FE		Y	Y		Y
Individual Controls			Y		Y
Adj. R^2	0.001	0.006	0.016	0.002	0.015
Observations	4,534	4,534	4,534	4,534	4,534

Notes. This table reports OLS estimates of engagement on the disclosure condition, with heteroskedasticity-robust standard errors in parentheses; base arm = the control group. The individual controls added in model (3) are graduate degree, inferred female, an indicator for names the dictionary cannot classify, year of first enrollment, an indicator for enrollment year being missing, an indicator for business programs, an indicator for humanities programs, and an indicator for field not on file, with technical programs the omitted category. Models (4) and (5) add the randomized wage and an indicator for no wage shown; model (4) carries no other controls and model (5) carries every block at once. The wage enters centred on the mean displayed wage of $22.50, so the no-wage indicator is the contrast against a posting showing that wage. * $p < .10$, ** $p < .05$, *** $p < .01$, two-tailed.

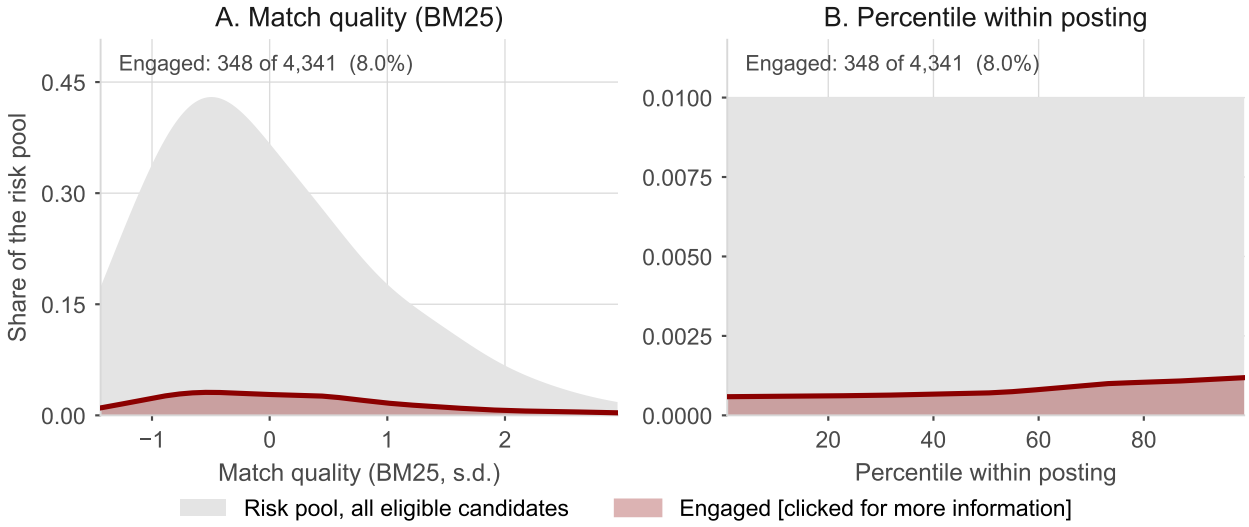
Table 6: The Composition Effect:
AI Disclosure × Match Quality

Dep. Var.:	*Engagement (click-through)*					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
AI matching	−0.024**	−0.024**	−0.022**	−0.023**	−0.024**	−0.024**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Match quality (BM25)	0.029***	0.033***	0.031***	0.023**	0.030***	0.021**
	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.009)
AI × Match quality	−0.028***	−0.028***	−0.028***	−0.022*	−0.026**	−0.020*
	(0.010)	(0.010)	(0.010)	(0.011)	(0.010)	(0.011)
				Profile Length	*Positions Listed*	*Seniority Index*
Rival attribute				0.023**	0.008	0.024**
				(0.010)	(0.008)	(0.010)
AI × Rival attribute				−0.015	−0.008	−0.018
				(0.013)	(0.010)	(0.012)
Job Posting FE		Y	Y	Y	Y	Y
Individual Controls			Y			
Adj. R^2	0.007	0.011	0.014	0.014	0.011	0.014
Observations	2,924	2,924	2,924	2,924	2,924	2,924

Note. OLS estimates on the 2,924 recommendations in the AI and human arms; the dependent variable is engagement (click-through). Match quality and the rival attributes are standardized, so coefficients read per standard deviation. Heteroskedasticity-robust standard errors in parentheses; human matching is the reference arm. Individual controls are the posted wage, graduate status, inferred gender and start year with their missingness indicators, together with posting and field-of-study fixed effects. Columns (4)–(6) each add the standardized rival attribute named directly above its coefficients, and that attribute's interaction with AI disclosure. Randomization inference for the main Section 4 contrasts is in Appendix Table B2. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

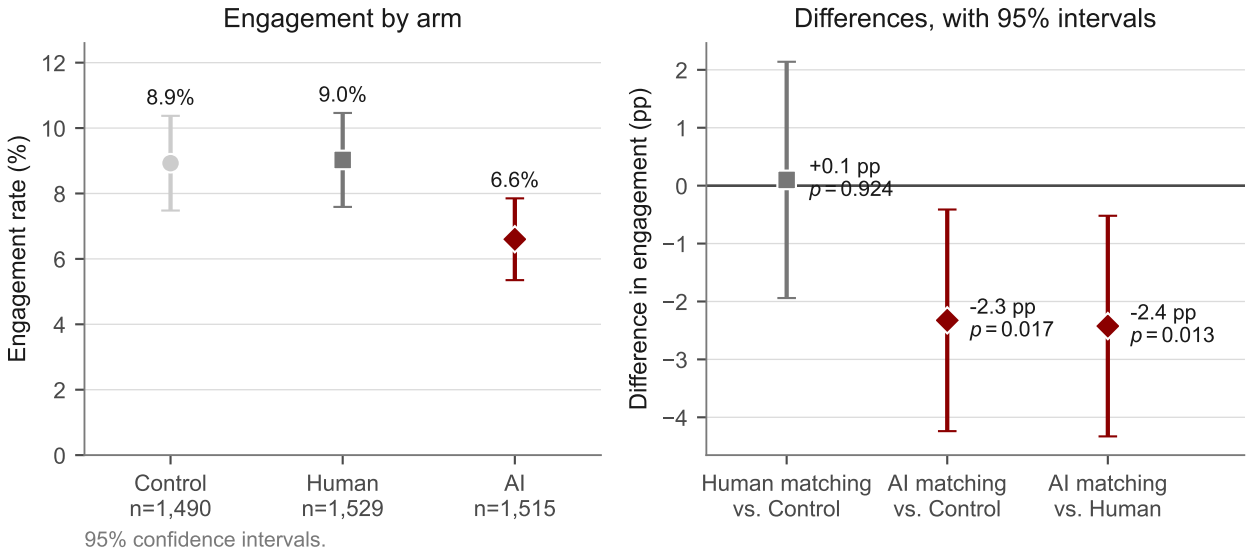
FIGURES

Figure 1: Just a Subset of the Overall Relevant Jobseeker Pool Engages With Postings



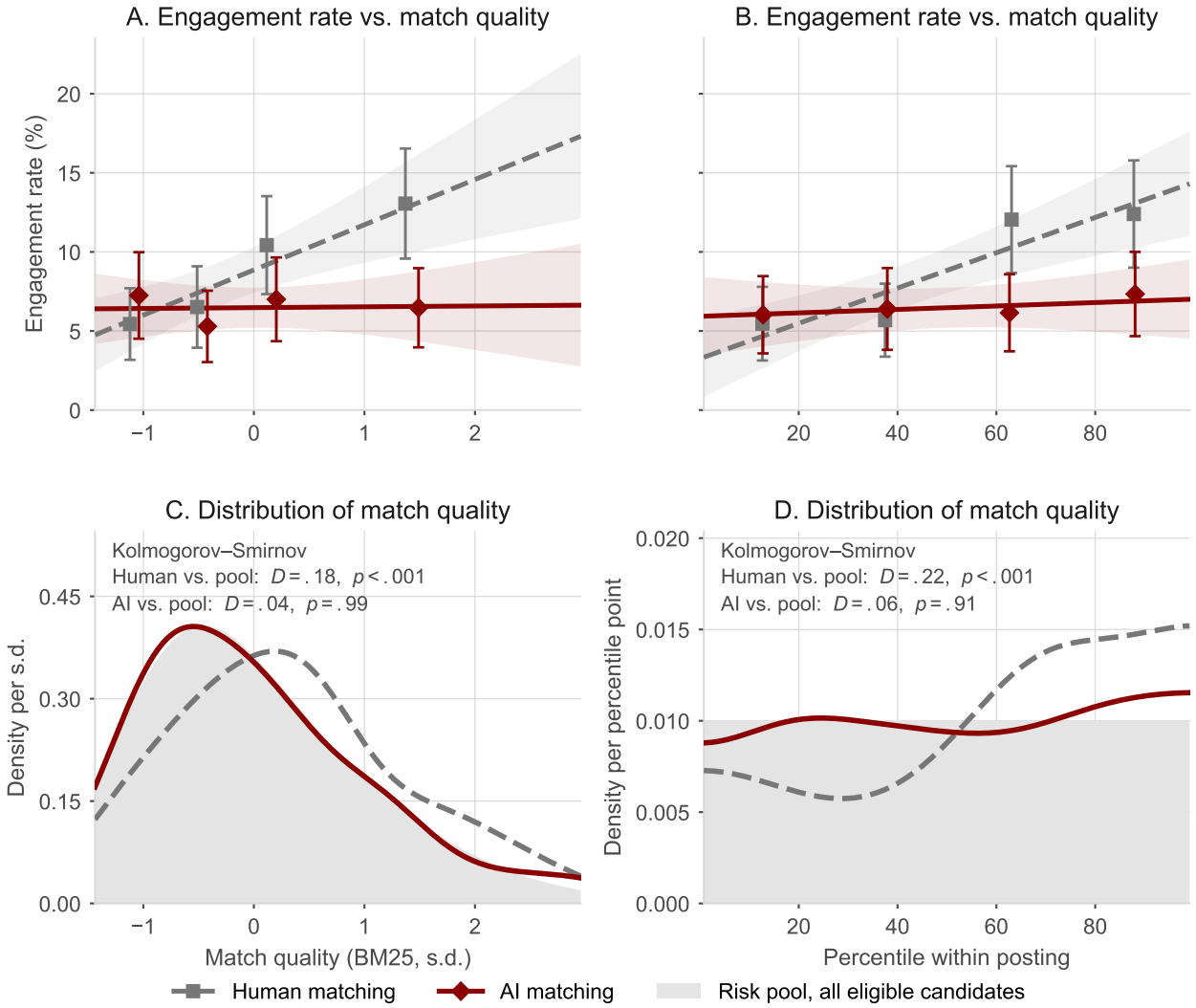
Note. All arms pooled; no treatment enters this figure. Of the 4,341 eligible candidates who received a recommendation and have a match-quality score, 348 engaged, or 8.0 percent. The shaded grey area is the distribution of match quality across the whole risk pool and integrates to one; the red area is the engaged candidates and integrates to 0.080, so the area under it is literally the share of the pool that engages. *Panel A* uses BM25 pooled across postings, *Panel B* each candidate's percentile within their own posting, where the pool is uniform by construction. Engagement is therefore both narrow and selective: the jobseekers who in fact engage are a small and better-matched slice of those the platform judged relevant.

Figure 2: AI Disclosure Reduces How Many Jobseekers Engage



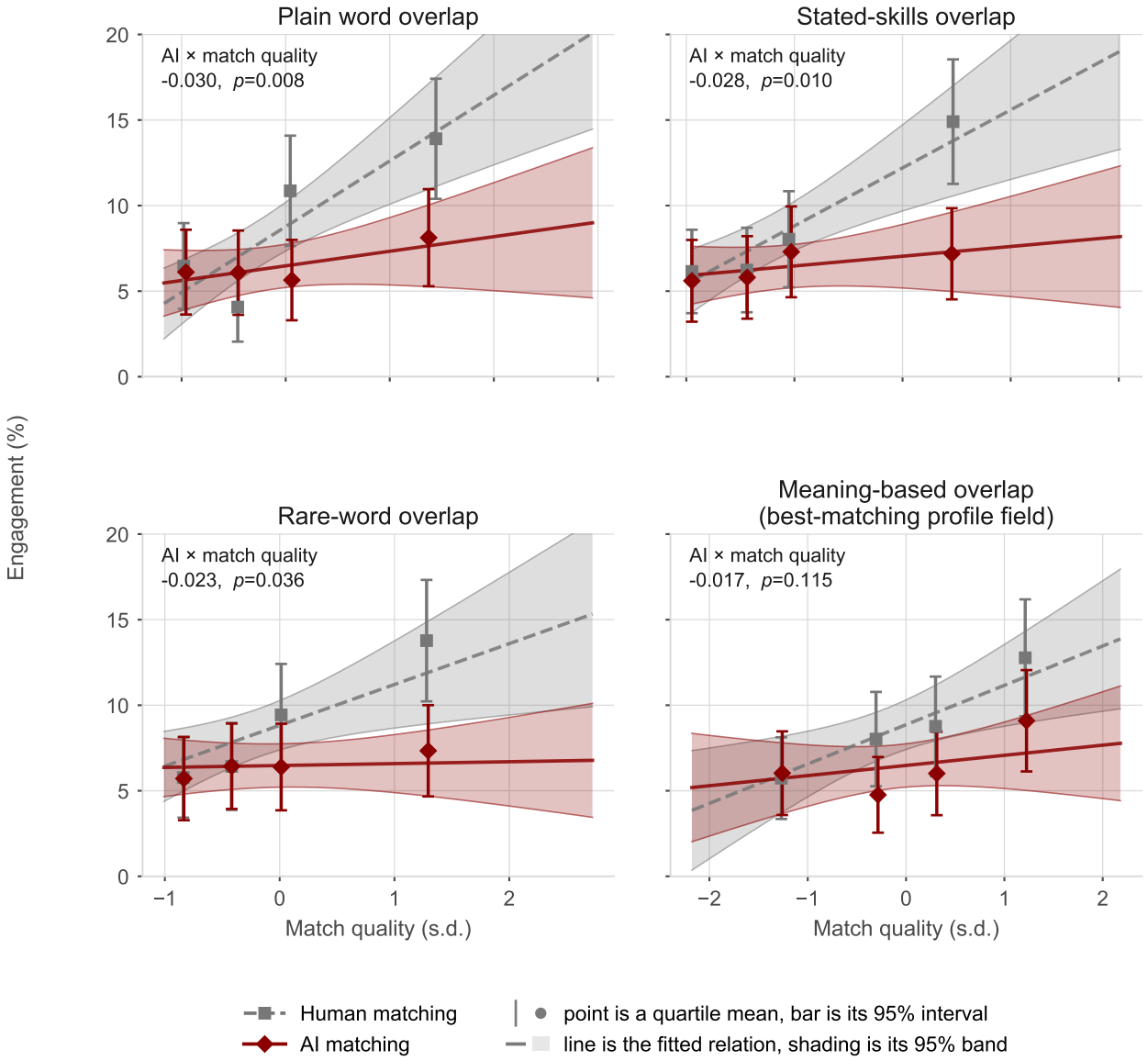
Note. Full sample ($N = 4{,}534$). *Left*: mean engagement (click-through rate) by arm. *Right*: the differences those means imply, with 95% confidence intervals. AI disclosure lowers engagement by 2.3 percentage points relative to the control group, a 26% reduction ($p = .017$; randomization inference $p = .021$). Human matching is indistinguishable from the control group (0.1 pp, $p = .92$). Intervals are 95% confidence intervals using heteroskedasticity-robust standard errors; treatment-effect inference uses the contrasts shown at right.

Figure 3: AI Disclosure Flattens the Match-Quality Gradient and Changes Who Engages



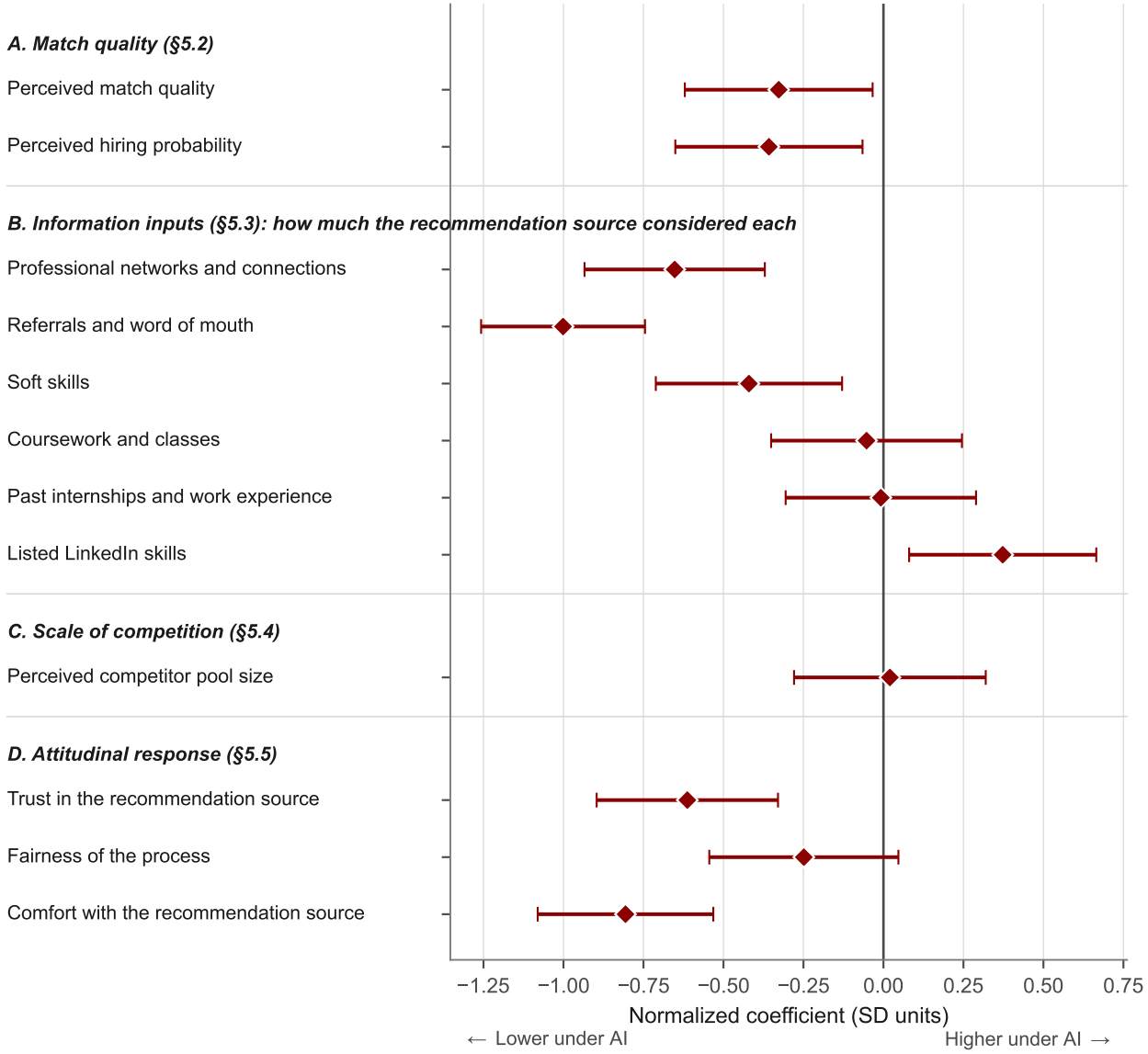
Note. The left column uses BM25 pooled across postings, in standard deviations; the right uses each candidate's percentile within their own posting, on its natural 0–100 scale, so they are compared only with others who faced the same job. *Panels A and B*: engagement rates on the 2,924 recommendations in the AI and human arms. Lines are fitted relations with 95% bands, from the model of Table 6, column (1), with heteroskedasticity-robust standard errors; points are quartile means with 95% intervals and a small horizontal offset between arms. The AI × match-quality interaction is −0.028 per standard deviation ($p = .006$) on the pooled measure, and −1.0 percentage points per ten percentile points ($p = .004$) within posting. *Panels C and D*: distributions of match quality for the full risk pool of 4,341 eligible candidates and for the candidates in each arm who engaged. Panel C is a kernel density estimate, bandwidth 0.43 standard deviations by Silverman's rule on the smallest group ($n = 94$) (Hansen, 2022). In panel D the risk pool is *uniform by construction* — a within-posting percentile must be — so it is drawn exactly rather than estimated, and the engaged curves are estimated with reflection at both bounds. Kolmogorov–Smirnov tests against the pool are reported in each panel. Panels A and C are truncated at the 99th percentile. Figure 4 repeats panel A for four alternative constructions, defined in Table 3.

Figure 4: The Same Qualitative Flattening Appears Across Alternative Measures of Match Quality



Note. Engagement against each alternative construction of match quality, separately by arm, on the 2,924 recommendations in the AI and human arms. The lines are the fitted relation with its 95% band, from the same model as Figure 3 applied to each measure; the points are quartile means with 95% intervals. The interaction annotated in each panel is from that same specification, with heteroskedasticity-robust standard errors. The measures are defined in Table 3.

Figure 5: Companion Lab Experiment:
AI–Human Differences in Beliefs and Attitudes



Note. The AI-versus-human difference on each item the companion lab experiment measured, grouped by the mechanism it speaks to and lettered A to D as in Section 2. Each estimate is a difference in means between the two randomly assigned arms, standardized by that item's pooled standard deviation, with 95% confidence intervals (lab vignette, $n \approx 176$). Twelve of the fourteen items appear here; the two stated intents are reported in Section 5.1. Question wordings and response scales are in Table A1, composites in Appendix Table B1. Under *information inputs* the three relational items are listed first and the three codified ones second. These are stated assessments under a hypothetical vignette, measured against a named human alternative rather than no source at all (Section 5.1).

APPENDIX A: COMPANION LAB VIGNETTE EXPERIMENT INSTRUMENT

This appendix reproduces the companion lab experiment exactly as administered (the instrument referenced in Section 5.1). The lab experiment was a two-arm, between-subjects study: respondents were randomly assigned to the **AI matching** arm or the **Human matching** arm. The two arms are identical except for the recommendation-source line in the vignette and the corresponding agent phrase piped into several questions (shown below in braces as `{AI arm / Human arm}`).

A.1 Vignette

Shown to all respondents (introduction):

> The following is an email sent to a student from the university. Please read it carefully — the questions that follow will ask for your impressions of the recommendation.

Recommendation email:

> **From:** [University] **To:** Alex **Subject:** A part-time opportunity recommended for you
> Dear Alex,
> Here is a part-time opportunity recommended for you.
>
> ---
>
> **RECOMMENDATION:** Research Assistant, Business & Marketing Strategy. Generate a B2B Marketing Strategy report.
> **RECOMMENDATION SOURCE:** {AI arm: *A.I. matching algorithm using publicly available Linkedin profiles.* / Human arm: *The employer generated a list of candidates by reviewing publicly available Linkedin profiles.*}
> **Estimated wage per hour:** $20
> [Go to Opportunity] — [University] —

(The construction of the lab experiment did not allow participants to return to the vignette after advancing.)

Two phrases are piped into later questions based on arm assignment: - **`{agent phrase}`** = "the A.I. matching algorithm" (AI arm) / "the employer" (Human arm) - **`{agent phrase, self}`** = "an A.I. matching algorithm" (AI arm) / "an employer" (Human arm)

A.2 Items and Labels

Table A1 lists the items in administration-block order; the six input items were randomized within their block. Each item is shown with the short label it carries in the paper's figures and tables. Group 1, the manipulation check on the recommendation's source, was asked first, on a separate page and from memory; background items (major, minor, school, year, transfer status, age, gender) and an optional open-ended comment followed the items below.

Table A1. Companion Lab Experiment: Item Dictionary

Label used in the paper's exhibits	*Question as administered*	*Scale*	*N*
2. Input beliefs			
Battery stem, six items in randomized order: "To what extent do you think {the A.I. matching algorithm / the employer} considered each of the following when making this recommendation?" (1 = not at all, 7 = a great deal)			
Professional networks and connections	"Alex's personal and professional network and connections"	1–7	177
Referrals and word of mouth	"Referrals or word-of-mouth recommendations about Alex"	1–7	177
Soft skills	"Alex's soft skills"	1–7	177
Coursework and classes	"Alex's coursework and classes"	1–7	177
Past internships and work experience	"Alex's past internships and work experience"	1–7	177
Listed LinkedIn skills	"Alex's listed skills on LinkedIn"	1–7	177
3. Match quality			
Perceived match quality	"How good a fit do you think this recommendation is for Alex?" (1 = not at all a good fit, 7 = a very good fit)	1–7	176
4. Pool size			
Perceived competitor pool size	"How many other candidates do you think received this recommendation?" (1 = very few, 7 = very many)	1–7	176
5. Hiring probability			
Perceived hiring probability	"If Alex applies for this position, how likely do you think Alex is to get it?" (slider)	0–100	177
6. Attitudes			
Trust in the recommendation source	"How much do you trust {the A.I. matching algorithm / the employer} to make a good recommendation?" (1 = not at all, 7 = completely)	1–7	177
Fairness of the process	"How fair do you think this recommendation process is?" (1 = not at all fair, 7 = very fair)	1–7	177
Comfort with the recommendation source	"How comfortable would you be with {an A.I. matching algorithm / an employer} making job recommendations for you personally?" (1 = not at all comfortable, 7 = very comfortable)	1–7	176
7. Stated engagement			
Click-through intent	"If you were Alex, how likely would you be to click through to learn more about this position?" (1 = definitely would not, 7 = definitely would)	1–7	176
Application interest	"If you were Alex, how interested would you be in applying for this position?" (1 = not at all interested, 7 = very interested)	1–7	177
Composites (not administered; constructed from the items above)			
Relational inputs, composite	Equally weighted mean of the z-scored network, referral and soft-skill items	z	177
Codified inputs, composite	Equally weighted mean of the z-scored coursework, internship and listed-skill items	z	177
Attitudes, composite	Equally weighted mean of the z-scored trust, fairness and comfort items	z	177

Notes. This table shows the substantive items the companion lab experiment administered, in administration-block order, with the six input items randomized within their block, together with the short label each carries in the paper's figures and tables. The estimated AI-versus-human difference is plotted in Figure 5 for the belief and attitude items, and reported in each item's own scale units in Appendix Table B1; the two stated-engagement items are reported in Section 5.1. This table is the dictionary those exhibits refer to and reports no estimates of its own. N is the number answering that item; 178 respondents were randomized (87 AI, 91 Human) and one or two skip individual items. Bracketed phrases were replaced by the respondent's assigned condition. The vignette text is in Section A.1; the administration order is the order of this table.

APPENDIX B: SUPPLEMENTARY TABLES

B.1 Correct-Recall Subsample

Table B1. Lab Experiment: Full Randomized Sample vs. Correct-Recall Subsample

	Full randomized sample		Correct-recall subsample	
Outcome	AI disclosure	N	AI disclosure	N
Panel A. Stated engagement				
Click-through intent (1–7)	-0.982^{***}	176	-1.357^{***}	126
	(0.245)		(0.280)	
Application interest (1–7)	-0.798^{***}	177	-1.022^{***}	127
	(0.236)		(0.290)	
Panel B. Match quality (Section 5.2)				
Perceived match quality (1–7)	-0.335^{**}	176	-0.440^{**}	126
	(0.152)		(0.186)	
Perceived hiring probability (0–100)	-7.907^{**}	177	-8.912^{**}	127
	(3.274)		(4.060)	
Panel C. Information inputs (Section 5.3)				
Relational-input composite (z)	-0.691^{***}	177	-0.986^{***}	127
	(0.110)		(0.127)	
Codified-input composite (z)	0.104	177	-0.146	127
	(0.115)		(0.127)	
Listed LinkedIn skills (1–7)	0.561^{**}	177	0.199	127
	(0.223)		(0.259)	
Panel D. Scale of competition (Section 5.4)				
Perceived competitor pool size (1–7)	0.026	176	0.208	126
	(0.193)		(0.225)	
Panel E. Attitudinal response (Section 5.5)				
Attitudinal composite (z)	-0.553^{***}	177	-0.742^{***}	127
	(0.125)		(0.146)	

Notes. This table reports the preregistered correct-recall specification beside the full randomized sample used in the paper. Recall was itself affected by treatment: 76 of 87 AI-arm respondents but only 52 of 91 in the Human arm recalled the source correctly. Conditioning on recall therefore breaks the original randomization, so the full sample is primary and the screened sample is reported as a sensitivity check (Montgomery et al., 2018). All main estimates are larger in the screened sample and retain their significance, which is what dropping respondents who did not register the manipulation would produce rather than evidence that the screened specification is the better one; perceived pool size remains null. The exception is listed LinkedIn skills: its estimate falls from 0.56 ($p = .012$) in the full sample to 0.20 ($p = .44$) under the recall screen. Section 5.3 therefore treats the full-sample increase cautiously: the screened estimate remains non-negative, but the evidence for an increase is no longer precise. Each entry is a separate OLS difference in means with heteroskedasticity-robust standard errors. The composites are equally weighted means of their z-scored items, so a composite's N counts respondents answering at least one of its three items. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$, two-tailed.

B.2 Randomization Inference

Treatment was assigned at random among the candidates recommended for a given posting, so the distribution of estimates under the hypothesis of no effect can be constructed from that assignment rather than assumed (Young, 2019). We reassign the arm labels among the candidates within a posting 4,000 times, recompute the statistic on each reassignment, and report the share of reassignments whose estimate is at least as large in absolute value as the realized one. Because the reassignment is over all three arms, which candidates fall in the pair being compared moves from draw to draw, as it would have under a different assignment. The procedure assumes nothing about the distribution of engagement, and it does not produce a standard error; it produces a p-value directly.

Table B2 reports randomization-inference results for the 14 main treatment and treatment-by-match-quality contrasts, next to the heteroskedasticity-robust p-value from the corresponding main-text estimate. The two never disagree about significance at conventional levels, and no p-value differs by more than .021. The largest discrepancies fall on contrasts that are far from any threshold of significance.

Table B2. Randomization Inference for Main Field-Experiment Contrasts

	Estimate	(s.e.)	Robust p	RI p	N
Level					
AI vs. control, arms only	−0.0233	(0.0098)	.017	.021	3,005
AI vs. control, + posting FE	−0.0236	(0.0097)	.015	.015	3,005
AI vs. control, + posting FE, controls	−0.0233	(0.0098)	.017	.016	3,005
AI vs. human matching	−0.0242	(0.0097)	.013	.014	3,044
Human matching vs. control	+0.0010	(0.0104)	.924	.918	3,019
Composition					
AI × match quality, vs. human	−0.0281	(0.0102)	.006	.010	2,924
+ posting FE	−0.0283	(0.0102)	.006	.006	2,924
+ full controls	−0.0279	(0.0104)	.007	.005	2,924
AI × match quality, vs. control	−0.0195	(0.0106)	.067	.066	2,868
Human × match quality, vs. control	+0.0086	(0.0116)	.462	.441	2,890
Alternative measures					
AI × plain word overlap	−0.0299	(0.0112)	.008	.009	2,924
AI × rare-word overlap	−0.0228	(0.0108)	.036	.037	2,924
AI × stated-skills overlap	−0.0282	(0.0109)	.010	.016	2,924
AI × meaning-based overlap	−0.0171	(0.0108)	.115	.106	2,924

Note. Each row is a contrast reported in Section 4, estimated by a linear probability model on the recommendation-level data. *Estimate* is the arm difference in engagement for the level panel and the arm × match-quality interaction for the other two, with heteroskedasticity-robust standard errors in parentheses. *Robust p* is from those standard errors; *RI p* is the share of 4,000 within-posting reassignments of the three arm labels whose estimate is at least as large in absolute value as the realized one. Match quality is standardized, so interactions read per standard deviation. Samples differ across rows because the level panel uses the full sample while the others require a match-quality score, and because each row uses only the two arms it compares.

APPENDIX C: THE RECOMMENDATION EMAIL AND ITS RANDOMIZED TEXT

Every jobseeker in a posting's risk set received one email of the form below, sent from a university address. The email was the only channel through which these positions were advertised, so the link at its center was the sole route to the posting. Angle brackets mark fields filled in per recipient; the university and platform names are replaced here for review. Three elements were randomized, and they are set out after the template.

> Dear ⟨first name⟩,
>
> Here is a part-time ⟨noun⟩ you can do in your spare time for a few weeks this semester.
>
> --
>
> Recommendation: ⟨one-line job description⟩
>
> --
>
> ⟨recommendation-source line⟩
>
> ⟨expected wage line⟩
>
> Go to ⟨platform⟩ Opportunity
>
> We are always experimenting and re-testing to improve our systems. We always appreciate whatever comments or suggestions you might have.
>
> - Your ⟨university⟩ ⟨platform⟩ Team!

The disclosure treatment. The recommendation-source line is the treatment. Jobseekers assigned to *AI matching* received the line "Recommendation source: A.I. matching algorithm using publicly available LinkedIn profiles." Jobseekers assigned to *human matching* received the line "Recommendation source: The employer generated a list of candidates by reviewing publicly available LinkedIn profiles." Jobseekers assigned to the *control* arm received no such line: the email named no source at all, and nothing else in it differed. The two treatment lines are equal in length to within a few words and differ in what they attribute the match to, not in how much they say.

The wage treatment. The expected wage line named an hourly rate drawn independently from \$16 to \$28 in \$2 increments. Jobseekers assigned to the wage control received no wage line.

A co-randomized noun. The noun in the first line was independently randomized among "opportunity," "gig," and "contest" for a separate study. It was randomized independently of the disclosure and wage arms, and we report it here for completeness.