# Artificial Intelligence and the Alleviation of Labor Market Mismatch: Evidence from Online Job Matching Data*

Hang Yu† Hongbo Li‡ Dandan Zhang§ Runbo Zhang‡ Qiang Li¶

This version: August 2026

**Abstract**

Using 1.62 million matched vacancy–applicant observations from an online recruitment platform, we refine an occupation-level large language model (LLM) exposure index, construct measures of vertical and horizontal mismatch, and apply difference-in-differences and event-study designs to identify the impact of LLM technologies on labor market mismatch. We show that overall mismatch increased after 2021. But following the technological shock associated with ChatGPT, vertical mismatch declined in occupations with higher LLM exposure, accompanied by more precise vacancy signals, higher skill requirements, and educational thresholds. These results provide vacancy–applicant evidence on how AI improves labor market matching efficiency.

**Keywords:** technological progress; labor market mismatch; LLM exposure index
**JEL:** J24, J62, O33

## 1 Introduction

Large language models (LLMs) are changing not only how much labor firms demand but what they ask of workers. As firms reorganize job tasks around the new technology, skill requirements rise and become more finely specified, and these demand-side signals feed back into workers' schooling and occupational choices. The question of how artificial intelligence affects employment is therefore shifting from a pure quantity question—displacement versus creation—to a question about the structure and efficiency of matching between workers and jobs. This paper asks a specific version of that question with direct welfare stakes:

*We thank Zhaopin.com for data support. This paper is a working-paper version, adapted for an international audience, of an article published in Chinese in the *China Economic Quarterly* (2026, 26(3): 874–894, DOI: 10.13821/j.cnki.ceq.2026.03.11). The LLM exposure index used here refines the index of Zhang et al. (2025) and is documented in the public release of the AI and Economics Laboratory at the National School of Development, Peking University (AI and Economics Laboratory, National School of Development, Peking University and Zhaopin, 2025). Financial support from the National Natural Science Foundation of China (Grant No. 72303009) is gratefully acknowledged. Dandan Zhang (ddzhang@nsd.pku.edu.cn) is the corresponding author of the Chinese publication. Correspondence regarding this working paper: Hang Yu (hangyu.economics@gmail.com).

†China Center for Economic Research and National School of Development, Peking University; Institute of South–South Cooperation and Development.

‡National School of Development, Peking University.

§China Center for Economic Research and National School of Development, Peking University.

¶Zhaopin Limited.

does LLM-driven technological change worsen or alleviate *education mismatch*—the misalignment between workers' schooling and the jobs they end up in?

China is a natural setting in which to ask it. The country's higher-education expansion keeps delivering record graduate cohorts just as LLM applications diffuse through firms, so "technology-driven restructuring of demand" and "scale-driven growth of supply" land on the labor market at the same time. Following the OECD's taxonomy, mismatch comprises qualification (education-level) mismatch, skill mismatch, and field-of-study mismatch (OECD, 2017); constrained by what recruitment data can measure, we focus on the first and third—which we call vertical and horizontal mismatch—and leave skill mismatch aside. Evidence on how AI affects these margins is scarce. Internet access has been shown to improve education–job matching (Xie et al., 2024), but LLMs restructure job tasks, skill demands, and recruitment signals far more deeply than generic information technology, and the direction of their effect on mismatch is an open empirical question.

Answering it requires observing mismatch where it forms—at the point of matching—rather than in the stock of existing employment. We use matched vacancy–applicant data from Zhaopin, a major Chinese online recruitment platform: 97,447 randomly sampled job postings from January 2021 to July 2025, the 1,623,430 applications they received, and the 1,439,854 distinct applicants behind them. The platform records employers' responses to applications, so we observe not only who applies where but which applications receive a *positive reply*—our proxy for a near-hire. We measure a "downward application" as one in which the applicant's education exceeds the posting's requirement, and a "cross-field application" as one in which the applicant's field of study falls outside the posting's occupation under the CIP–SOC crosswalk. Among applications that receive a positive reply, these become our measures of vertical and horizontal mismatch.

To identify the effect of LLM-driven technological change, we treat the release of ChatGPT in December 2022 as a salient technology shock and compare occupations with different baseline LLM exposure in difference-in-differences and event-study designs. Exposure is measured by an occupation-level LLM exposure index built from roughly 600,000 job postings from 2018–2020—before our matching sample begins, which guards against reverse causality from post-LLM job redesign. The index refines the measure of Zhang et al. (2025): following Hampole et al. (2025), task extraction is decomposed into four transparent steps, and postings are mapped to the O*NET task inventory with Sentence-BERT semantic matching rather than a single end-to-end model prompt.

Four findings emerge. First, education mismatch in China's labor market is high and rising. Between 2021 and 2025 the vertical mismatch share among near-hires rose from 52.0% to 64.9%, and the horizontal mismatch share rose from a low of 40.7% to a high of 49.3%. Notably, applicants' behavior did not drive this trend: the shares of downward and cross-field *applications* stayed roughly stable at 40–45%. The rise in mismatch reflects employer selection—mismatched applicants are increasingly represented among positive replies.

Second, after the ChatGPT shock, high-exposure occupations received more applications but returned fewer positive replies. A one-standard-deviation increase in occupational exposure raises applications per posting by 11.30 on the intensive margin while the positive-reply rate falls, indicating sharper competition

for exposed jobs.

Third—our main result—vertical mismatch fell in exposed occupations. After the shock, a one-standard-deviation increase in exposure reduces the downward-application share by 2.6 percentage points, and the vertical mismatch share among near-hires falls correspondingly. Event-study estimates show no differential pre-trends and a decline that steepens through late 2024 and early 2025. Cross-field applications and horizontal mismatch show no significant response.

Fourth, the mechanism runs through employer-side job redesign. After the shock, postings in exposed occupations list significantly more tasks and skills, and the skills they require become more complex and more specialized. Clearer, more demanding vacancy signals screen out underqualified and opportunistic applications, which is precisely where the decline in downward applications and vertical mismatch comes from. The adjustment is uneven across the job ladder—bachelor's-level postings hold stable while junior-college postings absorb displaced high-education applicants—and across workers: STEM-trained applicants hold their ground in exposed occupations, while non-STEM applicants and workers with prior experience in exposed occupations slide down the job ladder.

This paper speaks to three literatures. A first strand documents the causes and consequences of education mismatch—its high incidence, wage penalties, and scarring effects (Duncan and Hoffman, 1981; Verdugo and Verdugo, 1989; Hartog, 2000; McGuinness, 2006; Robst, 2007; Tsai, 2010; Somers et al., 2019; Li, 2016, 2024)—using job-analysis, realized-matches, and subjective-assessment measures (Eckaus, 1964; Verhaest and Omey, 2006; Bender and Heywood, 2011; Bender and Roche, 2013; Wolbers, 2003). In China, the rapid release of graduates from higher-education expansion (Li et al., 2023), together with individual heterogeneity (Frank, 1978; Green et al., 2007) and employers' wariness of overqualification (Kuhn and Shen, 2013), has amplified matching frictions. We contribute flow-based evidence measured at the point of matching rather than in the employment stock.

A second strand studies technology and labor markets. The skill-biased and routine-replacing character of earlier waves—computerization and industrial robots—is well established (Autor et al., 2003; Acemoglu and Autor, 2011; Autor and Dorn, 2013; Goos and Manning, 2007), including for China (Wang and Dong, 2020; Li et al., 2021; Yan et al., 2020; Chen et al., 2022; Wang et al., 2023). LLMs differ: their semantic understanding and cross-task transfer push the frontier from routine tasks into non-routine cognitive work (Brynjolfsson and McAfee, 2014; Brynjolfsson and Mitchell, 2017; Frey and Osborne, 2017; Eloundou et al., 2024), so their net effect turns on the balance of displacement and creation (Acemoglu and Restrepo, 2019; Bessen, 2019; Deming and Noray, 2020; Acemoglu et al., 2022). This literature has focused on employment quantities, job structure, and skill demand; direct evidence on mismatch is what we add.

A third strand measures AI exposure with task-based indices (Felten et al., 2021; Webb, 2020; Eloundou et al., 2024). Zhang et al. (2025) brought this approach to Chinese posting text; we refine their index with structured task decomposition and Sentence-BERT standardization, improving transparency, precision, and robustness.

Our contributions are threefold. First, using matched vacancy–applicant data, we identify the effect of an AI technology shock on education mismatch at the micro level of the matching process itself, extending a literature that has mostly studied occupations or aggregate job structure. Second, we deliver a refined,

more transparent occupation-level LLM exposure index for the Chinese labor market. Third, we show that generative AI can improve matching efficiency within exposed occupations through job specialization and clearer signaling even while aggregate mismatch keeps rising—evidence that employer responses, not just technology itself, shape AI's labor market consequences.

The paper proceeds as follows. Section 2 describes the data and the construction of the exposure index and mismatch measures. Section 3 presents trends, the identification strategy, and the main results. Section 4 examines mechanisms and heterogeneity. Section 5 discusses implications and limitations.

## 2 Data and Measurement

Our data come from Zhaopin, one of China's largest online recruitment platforms. Online postings track demand-side change in close to real time, which is a natural advantage for measuring technology exposure, and the platform also records applications and employer responses, which lets us observe matching as it happens. Online recruitment data over-represent urban, younger, more educated job seekers and mid-to-high-wage jobs (Kuhn and Shen, 2013; Zhang et al., 2025), so extrapolation to the entire labor market warrants caution.

### 2.1 Job postings and the LLM exposure index

We construct an occupation-level LLM exposure index for the Chinese labor market from a random sample of roughly 600,000 Zhaopin postings from 2018 to 2020. Because the index is built entirely from postings that predate the start of our matching sample in 2021, it is not contaminated by post-LLM job redesign. The index refines Zhang et al. (2025): following Hampole et al. (2025) and adapting to the structure of posting text, task extraction is decomposed into four steps, each leaving inspectable output, which makes the construction more transparent and traceable.[1]

### 2.2 Matched vacancy–applicant data

We build the matching sample from Zhaopin data covering January 2021 to July 2025. Sampling is from the demand side: each year we randomly draw about 20,000 postings (the "posting sample"), extract every application those postings received in the posting year (the "application sample"), and identify the applicants behind them (the "applicant sample"). The data record each posting's date, occupation, industry, and location, and each applicant's age, gender, education, field of study, and previous job. The platform flags employer–applicant interactions that indicate mutual interest; we treat this *positive reply* as a proxy for a near-hire. To keep application histories complete, we retain postings published in January–October of each year (January–May for 2025) and match applications within two months of posting, and we keep full-time

[1] The construction, and its comparison with Zhang et al. (2025), is documented in the technical release of the AI and Economics Laboratory at the National School of Development, Peking University, and Zhaopin (AI and Economics Laboratory, National School of Development, Peking University and Zhaopin, 2025). Appendix B summarizes the procedure, and Appendix C reports the exposure of each occupation. The 2018–2020 posting sample is identical to that of Zhang et al. (2025), which keeps the two indices comparable.

postings only. The final sample contains 97,447 postings, 1,623,430 applications, and 1,439,854 distinct applicants; over 90% of applications come from applicants with junior college education or above. The average positive-reply rate over the sample period is about 19.4%. Roughly half of postings receive no applications; postings that do receive applications have significantly higher LLM exposure, pay, and education requirements. Appendix D details the sample construction and reports descriptive statistics.

### 2.3 Measuring education mismatch from matched data

**Standardizing occupations and fields of study.** To compare posting requirements with applicant credentials, we map both sides into international classification systems: postings' occupations into the Standard Occupational Classification (SOC), and applicants' fields of study into the Classification of Instructional Programs (CIP). The posting sample spans nearly a thousand platform occupation categories, which we map to SOC occupations by combining LLM-assisted matching with expert review; the applicant sample contains about 44,000 self-reported majors, which we map to CIP fields using Sentence-BERT semantic similarity plus expert review. Field-of-study analyses focus on applicants with junior college education or above—over 90% of the sample, and the group whose field information is most reliable. Appendix F describes the SOC and CIP systems and the matching procedures in detail, and Appendix H details the field-of-study standardization.

**Vertical and horizontal mismatch.** We start from application behavior. A *downward application* is one in which the applicant's education level exceeds the posting's requirement, with education graded on five levels: no requirement or junior high school and below (1); senior high school, secondary vocational, or technical school (2); junior college (3); bachelor's degree (4); graduate degree (5). A *cross-field application* is one in which the applicant's field of study does not match the posting's occupation. Following the literature (Manuel and Plesca, 2020; Manuel, 2024), we take the CIP–SOC crosswalk maintained by the BLS and NCES as the reference: applications outside the crosswalk are cross-field. To limit borderline cases, we aggregate SOC occupations to minor groups (98) and CIP fields to major groups (48) before applying the crosswalk.[2] Because applicants with junior college education or above are the relevant population for field-of-study analysis, cross-field applications and horizontal mismatch are computed on applications from that group (the "junior-college-and-above subsample"), with no restriction on the posting's required education.

Mismatch properly refers to realized matches, which recruitment data cannot observe directly. We therefore measure mismatch among near-hires: within applications that received a positive reply, a downward application constitutes *vertical mismatch* and a cross-field application constitutes *horizontal mismatch*.[3]

In the application sample, 2.0% of records lack the applicant's education, and in the junior-college-and-above subsample, 1.6% lack usable field-of-study information (missing, or below the 0.5 semantic-

[2]The aggregation level affects measured horizontal mismatch; Appendix G reports average horizontal mismatch under alternative aggregations.

[3]Upward applications are conceptually also vertical mismatch, but applications from applicants whose education falls below the posting requirement almost never receive positive replies, so we do not analyze them separately.

similarity threshold to any CIP field).[4] These records differ systematically from the rest of the sample, so dropping them could bias the results; we instead impute their downward-application and cross-field status with machine-learning predictions based on the records' remaining information. Appendix I reports details.

# 3 LLM Exposure and Labor Market Mismatch

## 3.1 Trends in application behavior and mismatch

Figure 1 tracks application behavior and mismatch from 2021 to 2025. Throughout the period, roughly 40–45% of applications are downward applications and about 40% of applications from the junior-college-and-above subsample are cross-field—both fairly stable. Measured among near-hires, however, mismatch rose sharply: the vertical mismatch share climbed from 52.0% in 2021 to 64.9% in 2025, and the horizontal mismatch share rose from a low of 40.7% to a high of 49.3%. These levels exceed the 20–40% mismatch rates typical of survey-based studies (Duncan and Hoffman, 1981; Verdugo and Verdugo, 1989; Li et al., 2023), which measure the employment stock; our data measure the flow of near-hires at the recruitment stage, so levels are not directly comparable, but the series is internally consistent over time.[5]

That mismatch rises while application behavior stays flat pins the trend on employer selection: vertical mismatch runs well above the downward-application share, and horizontal mismatch above the cross-field

[4] Appendix H discusses the choice of threshold.

[5] Online recruitment data do not cover the full labor market. Zhaopin's applicants skew young and mid-to-highly educated—groups more prone to mismatch—so the levels in Figure 1 are likely upper bounds for the labor market as a whole.

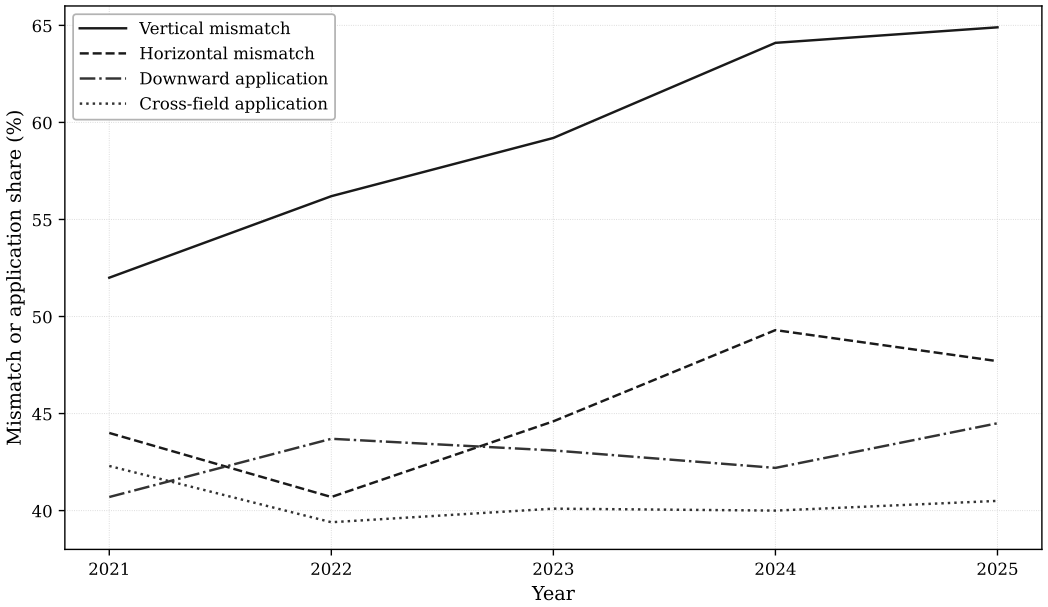



**Figure 1:** Application behavior and labor market mismatch, 2021–2025

*Notes:* Based on the matched vacancy–applicant data. Downward and cross-field application shares are computed over applications; vertical and horizontal mismatch shares are computed over applications receiving a positive reply (near-hires). Cross-field and horizontal series use the junior-college-and-above subsample. The figure is redrawn in English from the published version; underlying values are identical.

share, so employers increasingly favor over-educated and cross-field applicants in screening. The preference for downward applicants is especially pronounced—an intensifying "education premium" in screening that crowds out well-matched applicants and pushes vertical mismatch up. Rising vertical mismatch means part of the highly educated workforce cannot deploy its schooling advantage—a waste of human capital and a source of employment pressure and low job satisfaction; rising horizontal mismatch signals a widening gap between the structure of educational programs and industry demand. Appendix J reports mismatch patterns across occupation groups and applicant types.

### 3.2 Occupation-level mismatch changes and exposure

Figure 2 plots, for each SOC broad occupation, the change in its average mismatch share after the release of ChatGPT (December 2022–July 2025) relative to before (January 2021–November 2022) against its LLM exposure, dropping occupations with few observations. For vertical mismatch (left panel), more exposed occupations saw markedly smaller increases—the most exposed saw declines: the weighted least squares fit has slope $-17.48$ (SE 2.98). For horizontal mismatch (right panel) there is no comparable gradient (slope 0.09, SE 2.25). The next subsection subjects this pattern to a formal design.

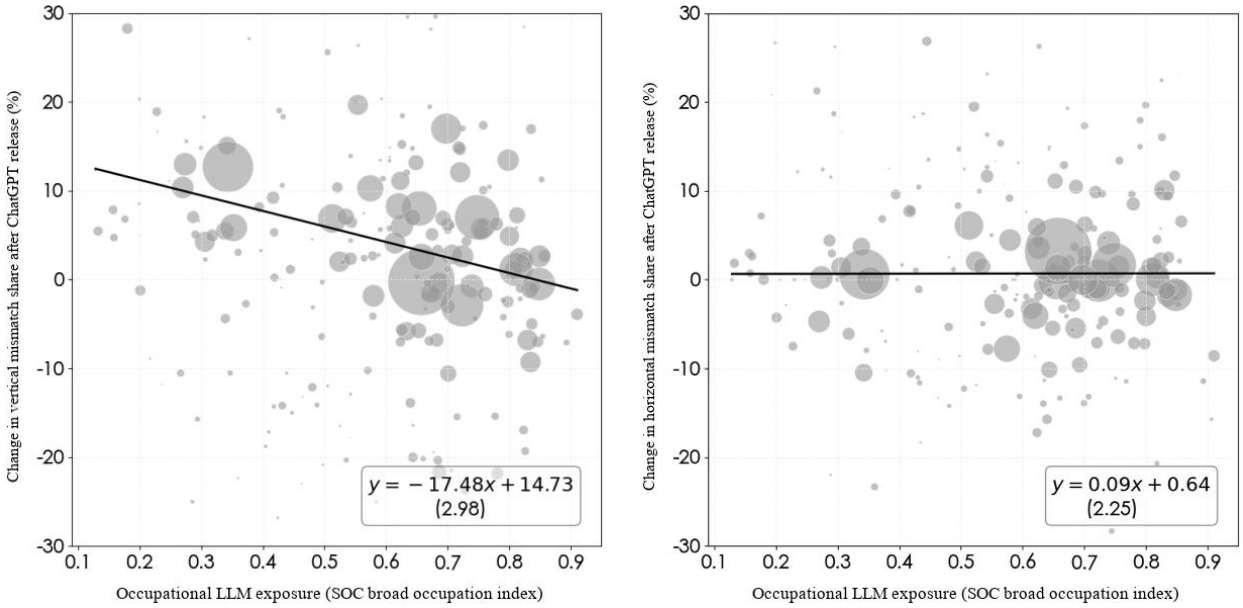



**Figure 2:** Occupation-level mismatch changes and LLM exposure

*Notes:* Left panel: vertical mismatch; right panel: horizontal mismatch. Each point is a SOC broad occupation (6-digit); point size is the occupation's share of near-hires before the ChatGPT release. The horizontal axis is the occupation's LLM exposure; the vertical axis is the difference between the occupation's average mismatch share after the release (December 2022–July 2025) and before it (January 2021–November 2022). The fitted line and its slope (standard error in parentheses) are from weighted least squares.

### 3.3 Empirical strategy

We take the posting as the unit of analysis and identify the causal effect of the LLM shock on mismatch with a difference-in-differences design, using the release of ChatGPT as the arrival of the shock and occupations' baseline LLM exposure as treatment intensity:

$$Y_{ijct} = \alpha + \beta \, \text{ExpOccu}_j \times \text{Post}_t + \gamma X_i + \mu_{ct} + \lambda_j + \varepsilon_{ijct}, \tag{1}$$

where $i$ indexes postings, $j$ the posting's SOC broad occupation, $c$ the city, and $t$ the year-month. We study two families of outcomes $Y_{ijct}$: application and positive-reply outcomes—whether the posting receives any application, the number of applications, and the positive-reply rate—and mismatch outcomes—the posting's downward-application share, vertical mismatch share, cross-field application share, and horizontal mismatch share. $\text{ExpOccu}_j$ is the occupation's LLM exposure index, Z-score standardized for interpretability. $\text{Post}_t$ equals one from December 2022 onward. $X_i$ is a vector of posting-level controls: number of planned hires, required education, monthly salary, firm ownership type, and firm size. $\mu_{ct}$ are month-by-city fixed effects, absorbing region-specific labor market trends including any platform strategy differences across regions; $\lambda_j$ are occupation fixed effects. Standard errors are clustered at the SOC broad occupation level, the level of the treatment variable.

To trace dynamics we also estimate the event-study version,

$$Y_{ijct} = \sum_{\tau \neq -1} \beta_\tau \, \mathbb{I}(t = \tau) \times \text{ExpOccu}_j + \gamma X_i + \mu_{ct} + \lambda_j + \varepsilon_{ijct}, \tag{2}$$

aggregating time to quarters for stability, with 2022Q4—the quarter of the ChatGPT release—as period 0.

### 3.4 Main results

Table 1 reports the effect of the shock on applications and positive replies. Column (1) shows no significant effect on the extensive margin—the probability that a posting receives any application. Column (2) shows that exposed occupations attracted more applications on the intensive margin: a one-standard-deviation increase in exposure raises applications per posting (among postings with applications) by 11.30 on average. More applications mean stiffer competition, and column (3) shows the positive-reply rate in exposed occupations falls significantly after the shock.

Table 2 contains the main results. Column (1): after the shock, more exposed occupations attract a smaller share of downward applications—one standard deviation of exposure reduces the downward-application share by 2.6 percentage points. Column (2) uses the sample of postings with positive replies, with the vertical mismatch share among those near-hires as the outcome: the shock lowers vertical mismatch in exposed occupations as well. Relative to less exposed occupations, exposed postings attract fewer downward applications *and* convert fewer of them into near-hires; Section 4 unpacks why. Columns (3) and (4) run the same designs for cross-field applications and horizontal mismatch and find no significant effect.

Figure 3 plots the event-study coefficients $\beta_\tau$ corresponding to columns (1)–(4) of Table 2. Before the

**Table 1:** Effects of LLM technology on market applications and replies

| Sample | All postings | Postings with applications | |
|---|---|---|---|
| Outcome | Any application (1) | Number of applications (2) | Positive-reply rate (3) |
| Post-ChatGPT × LLM exposure index | 0.015 | 11.30*** | −0.0300*** |
| | (0.0103) | (2.069) | (0.00613) |
| Posting controls | Yes | Yes | Yes |
| Month-by-city fixed effects | Yes | Yes | Yes |
| Occupation fixed effects | Yes | Yes | Yes |
| Observations | 91,897 | 46,462 | 46,462 |
| Adjusted $R^2$ | 0.482 | 0.157 | 0.334 |
| Mean of dependent variable | 0.517 | 33.274 | 0.343 |

*Notes:* "Posting controls" include the posting's number of planned hires, required education, monthly salary, firm ownership type, and firm size; column (3) additionally controls for the posting's number of applications. Standard errors in parentheses, clustered at the occupation level (SOC broad occupation, 6-digit). ***, **, * denote significance at the 1%, 5%, and 10% levels.

**Table 2:** Effects of LLM technology on labor market mismatch

| Posting sample | With applications | With positive replies | With applications# | With positive replies# |
|---|---|---|---|---|
| Outcome | Downward-application share (1) | Vertical-mismatch share (2) | Cross-field-application share (3) | Horizontal-mismatch share (4) |
| Post-ChatGPT × LLM exposure index | −0.0263*** | −0.0230*** | −0.00214 | −0.000123 |
| | (0.00589) | (0.00572) | (0.00344) | (0.00487) |
| Posting controls | Yes | Yes | Yes | Yes |
| Month-by-city fixed effects | Yes | Yes | Yes | Yes |
| Occupation fixed effects | Yes | Yes | Yes | Yes |
| Observations | 46,462 | 31,096 | 42,262 | 26,999 |
| Adjusted $R^2$ | 0.721 | 0.63 | 0.603 | 0.524 |
| Mean of dependent variable | 0.545 | 0.592 | 0.496 | 0.505 |

*Notes:* # indicates that columns (3)–(4) include only postings whose required education is junior college or above. "Posting controls" include the number of planned hires, required education, monthly salary, firm ownership type, firm size, and number of applications. Standard errors in parentheses, clustered at the occupation level (SOC broad occupation, 6-digit). ***, **, * denote significance at the 1%, 5%, and 10% levels.

shock, high- and low-exposure occupations show no significant differences or trends in downward applications or vertical mismatch (coefficients near zero); after 2022Q4 the interaction coefficients fall visibly, most sharply from late 2024 into early 2025.

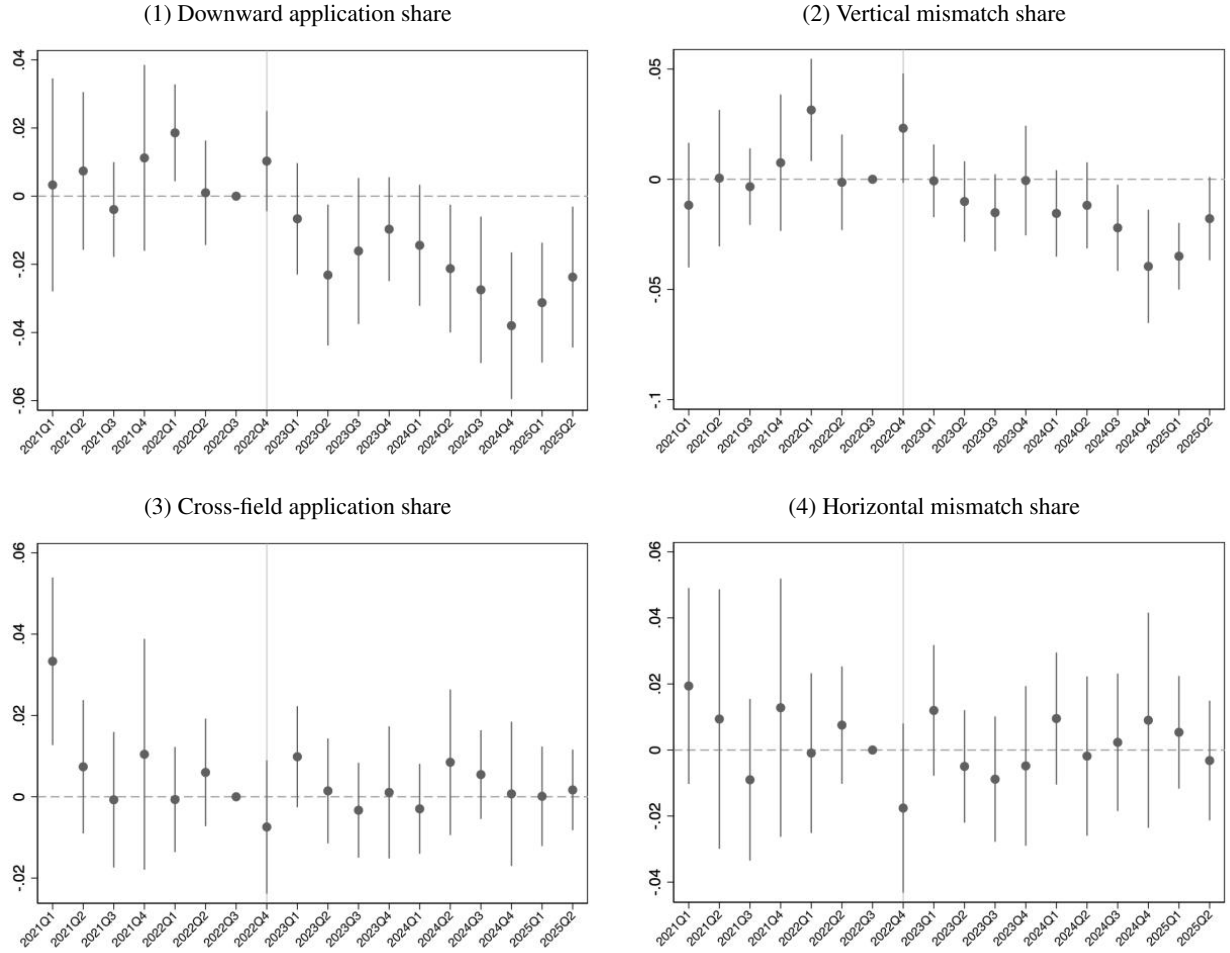



**Figure 3:** Dynamic effects of the LLM shock on mismatch

*Notes:* Each panel plots the event-study interaction coefficients from equation (2), with occupation and year-quarter fixed effects and posting-level controls (required education, salary, and other posting characteristics). Standard errors are clustered at the occupation level; bars are 95% confidence intervals. Panels (1)–(4) correspond to columns (1)–(4) of Table 2.

# 4 Mechanisms and Heterogeneity

Rapid AI progress confronts firms with task reorganization and skill substitution at once. For highly exposed occupations the shock moves both the boundary of automatable tasks and the internal skill structure of jobs, changing how firms write postings and how workers apply. This section explains the coexistence of rising aggregate mismatch with falling vertical mismatch in exposed occupations, and traces the micro mechanism underneath.

## 4.1 Demand side: task and skill specialization in exposed postings

The shock operates first on firms' production organization and stated demand. As generative AI diffuses, the task boundaries and skill demands of exposed occupations restructure systematically, making job content more complex and more specialized. We extract tasks from posting text and map them to the O*NET task

inventory, extract skills with an LLM (prompts in Appendix K) and map them to the Lightcast taxonomy by semantic matching, and, following Tacchella et al. (2012) and Aufiero et al. (2024), use the EFC algorithm to measure skill complexity and specialization. Table 3 shows that postings in exposed occupations adjusted significantly after the shock: task and skill counts rise—job content becomes more varied and the required skill bundle broader—while skill complexity rises, reflecting stronger complementarities among skills, and skill specialization rises, meaning skills shift from generic toward domain-specific and the job "profile" sharpens.

**Table 3:** Proactive adjustment of firms' job postings

| Dependent variable | Number of job tasks (1) | Number of skills (2) | Rank of highest skill complexity (3) |
|---|---|---|---|
| Post-ChatGPT × LLM exposure index | 0.0710** | 0.0812* | −2.013** |
| | (0.0344) | (0.0468) | (1.016) |
| Month-by-city fixed effects | Yes | Yes | Yes |
| Occupation fixed effects | Yes | Yes | Yes |
| Observations | 46,462 | 45,390 | 45,053 |
| Adjusted $R^2$ | 0.256 | 0.506 | 0.346 |
| Mean of dependent variable | 3.699 | 9.532 | 271.862 |

*Notes:* The regressions control for the posting's number of planned hires, required education, monthly salary, firm ownership type, firm size, and number of applications, as well as fixed effects at the year-month and city levels. Standard errors in parentheses, clustered at the occupation level (SOC broad occupation, 6-digit). ***, **, * denote significance at the 1%, 5%, and 10% levels.

In matching terms, restructured task and skill requirements raise the clarity of vacancy signals and the stringency of screening. When information is coarse, ill-matched applicants still enter the applicant pool; once skill requirements become explicit and specialized, those applications fall away—which is exactly where the declines in downward applications and vertical mismatch in exposed occupations come from.

## 4.2 Heterogeneity across the job ladder

Specialization changes not only skill demands within exposed occupations but the accessibility and appeal of jobs at different rungs. We examine the shock's effects by the posting's required education and salary. Table 4 shows that bachelor's-level postings kept stable downward-application and vertical mismatch shares—sharper skill signals stabilized matching at the top—while junior-college postings saw both rise significantly: as entry to bachelor's-level jobs moved up, highly educated applicants concentrated on middle-skill jobs, a "middle squeeze." Postings below junior college saw mismatch fall clearly, as highly educated applicants lost interest in low-skill jobs. The salary dimension agrees: downward applications and mismatch fell significantly in low-wage postings while high-wage postings changed little or ticked up. In sum, the shock pushed skill thresholds up at the top, intensified competition in the middle, and drained the bottom—differentiated mismatch adjustment along the job ladder.

**Table 4:** Regression results by the posting's required education and salary

| Grouping | By required education | | | | | | By posting salary | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sample | Bachelor's postings | | Junior college postings | | Below junior college | | High-salary postings | | Low-salary postings | |
| Outcome | Downward-app. share | Vertical-mism. share | Downward-app. share | Vertical-mism. share | Downward-app. share | Vertical-mism. share | Downward-app. share | Vertical-mism. share | Downward-app. share | Vertical-mism. share |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Post-ChatGPT × | 0.00595 | −0.00990 | 0.0239*** | 0.0242* | −0.0332*** | −0.0358*** | −0.0124* | −0.00549 | −0.0230*** | −0.0231*** |
| LLM exposure index | (0.00773) | (0.0162) | (0.00919) | (0.0131) | (0.00630) | (0.00557) | (0.00714) | (0.0101) | (0.00623) | (0.00666) |
| Posting controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Month-by-city fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Occupation fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 9,197 | 4,844 | 14,352 | 9,040 | 18,923 | 14,300 | 14,673 | 8,836 | 29,913 | 20,804 |
| $R^2$ | 0.337 | 0.309 | 0.306 | 0.311 | 0.243 | 0.238 | 0.757 | 0.676 | 0.708 | 0.627 |
| Mean of dependent variable | 0.111 | 0.114 | 0.452 | 0.467 | 0.860 | 0.866 | 0.453 | 0.514 | 0.590 | 0.624 |

*Notes:* Odd columns restrict the sample to job postings with applications; even columns restrict the sample to job postings with positive replies. "Downward-app. share" is the downward-application share and "vertical-mism. share" the vertical-mismatch share of the posting. "Posting controls" include the number of planned hires, required education, monthly salary, firm ownership type, firm size, and number of applications. Standard errors in parentheses, clustered at the occupation level (SOC broad occupation, 6-digit). ***, **, * denote significance at the 1%, 5%, and 10% levels.

## 4.3 Supply side: skill comparative advantage and transferability

We turn to how applicants with different human capital respond, moving to the application level for finer resolution:

$$Y_{pijct} = \alpha + \beta\,\text{ExpCurOccu}_j \times \text{Post}_t + \gamma X_i + \theta Z_p + \lambda_j + \mu_{ct} + \varepsilon_{pijct}, \tag{3}$$

where $p$ indexes applicants, $Z_p$ are applicant-level controls, and the outcome is whether the application is downward—or, within positive-reply applications, whether it constitutes vertical mismatch; other terms are as in equation (1).

Table 5 examines heterogeneity by field of study and work history. STEM applicants show no significant rise in downward applications or vertical mismatch in exposed occupations—some groups show lower mismatch—indicating that technically trained applicants adapt to the new skill structure and hold their position on high-skill tracks. Non-STEM applicants' downward applications and vertical mismatch rise significantly: as skill requirements climb, their competitiveness in high-skill jobs erodes and they shift toward middle-skill jobs, intensifying mid-ladder competition. Work history matters symmetrically: applicants with prior experience in highly exposed occupations show higher downward-application and vertical-mismatch rates after the shock—the task restructuring hit their existing skills directly, pushing them down the ladder—while applicants without such experience adjusted mildly.

Overall, the shock works through both sides of the market: on demand, specialization sharpens vacancy signals and lowers within-occupation mismatch; on supply, workers with different skills and experience face different adjustment paths, reinforcing the structural reshaping of mismatch.

**Table 5:** Regression results by applicant field of study and work experience (application level)

| Grouping | By applicant field of study | | | | By applicant work experience | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sample | STEM applicants | | Non-STEM applicants | | High-exposure work experience | | Low-exposure work experience | | No work experience | |
| Outcome | Downward application | Vertical mismatch | Downward application | Vertical mismatch | Downward application | Vertical mismatch | Downward application | Vertical mismatch | Downward application | Vertical mismatch |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Post-ChatGPT × LLM exposure index | 0.000576 | −0.00363 | 0.00865** | 0.0139*** | 0.00775** | 0.00996*** | 0.00263 | 0.00541 | −0.000607 | 0.00525 |
| | (0.00315) | (0.00380) | (0.00343) | (0.00353) | (0.00383) | (0.00320) | (0.00312) | (0.00388) | (0.00421) | (0.00590) |
| Posting controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Applicant controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Month-by-city fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Occupation fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 459,380 | 78,756 | 967,768 | 167,777 | 1,235,891 | 208,671 | 232,448 | 61,944 | 94,735 | 22,690 |
| $R^2$ | 0.779 | 0.792 | 0.784 | 0.768 | 0.760 | 0.751 | 0.751 | 0.772 | 0.792 | 0.767 |
| Mean of dependent variable | 0.392 | 0.541 | 0.430 | 0.599 | 0.400 | 0.552 | 0.515 | 0.646 | 0.501 | 0.676 |

*Notes:* The unit of observation is the vacancy–applicant application; the outcomes are indicators for whether the application is a downward application and whether it constitutes a vertical mismatch. Even columns restrict the sample to applications that received a positive reply. "Posting controls" include the number of planned hires, required education, monthly salary, firm ownership type, firm size, and number of applications. "Applicant controls" include the applicant's age, gender, education level, and tier of graduating institution. Standard errors in parentheses, clustered at the occupation level (SOC broad occupation, 6-digit). ***, **, * denote significance at the 1%, 5%, and 10% levels.

# 5 Discussion

LLM-driven technological change has not simply worsened mismatch across the board. By restructuring job tasks and skills, it made vacancy signals in exposed occupations clearer and reduced vertical mismatch within them—even as aggregate mismatch kept rising with marked stratification across job tiers and worker groups. Several implications follow. Monitoring systems built on posting text—tracking occupations' exposure, skill requirements, and mismatch in near real time—could identify high-mismatch occupations and at-risk groups before structural pressure accumulates. Education policy faces a matching problem rather than a quantity problem: the binding question is no longer whether enough educated workers are produced but whether their skills fit redesigned jobs, which argues for tighter articulation between programs and job skill demand, more training in AI-complementary and transferable capabilities, and application-oriented junior college education that does not merely absorb displaced degree-holders. Employment support is best aimed at the groups our estimates identify as vulnerable—non-STEM workers and those whose experience is concentrated in exposed occupations—with retraining tied to concrete occupational transitions rather than generic digital literacy. And employers themselves reduce mismatch when they post clearer task, skill, and career information: screening on demonstrated skills rather than degree thresholds alone would preserve entry for capable applicants without formal credentials.

Two limitations qualify the analysis. First, demand- and supply-side adjustments interact in general equilibrium—applicants' behavioral responses to the shock are themselves shaped by the demand restructuring it induces—so the two channels cannot be fully separated. Second, the shock will also operate slowly through human capital investment and skill formation; our flow data on new postings cannot speak to those long-run adjustments. Longer panels and richer identification are natural next steps.

## References


Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, volume 4B, chapter 12, pages 1043–1171. Elsevier.

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, 40(S1):S293–S340.

Acemoglu, D. and Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, 33(2):3–30.

AI and Economics Laboratory, National School of Development, Peking University and Zhaopin (2025). Methodology and data explanation for China's "artificial intelligence–large language model technology" exposure index. Technical documentation v2025.09, Peking University. Available at `https://nsd.pku.edu.cn/xzyj/kyfb/zsfb/ai/542352.htm`.

Aufiero, S., De Marzo, G., Sbardella, A., and Zaccaria, A. (2024). Mapping job fitness and skill coherence into wages: An economic complexity analysis. *Scientific Reports*, 14(1):11752.

Autor, D. (2015). Why are there still so many jobs? the history and future of workplace automation. *Journal of Economic Perspectives*, 29(3):3–30.

Autor, D. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review*, 103(5):1553–1597.

Autor, D., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4):1279–1333.

Béduwé, C. and Giret, J.-F. (2011). Mismatch of vocational graduates: What penalty on French labour market? *Journal of Vocational Behavior*, 78(1):68–79.

Bender, K. A. and Heywood, J. S. (2011). Educational mismatch and the careers of scientists. *Education Economics*, 19(3):253–274.

Bender, K. A. and Roche, K. (2013). Educational mismatch and self-employment. *Economics of Education Review*, 34:85–95.

Bessen, J. (2019). Automation and jobs: When technology boosts employment. *Economic Policy*, 34(100):589–626.

Brynjolfsson, E. and McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company, New York.

Brynjolfsson, E. and Mitchell, T. (2017). What can machine learning do? workforce implications. *Science*, 358(6370):1530–1534.

Chen, Y., Zhang, J., and Zhou, Y. (2022). Industrial robots and the spatial allocation of labor. *Economic Research Journal*, (1):172–188. (in Chinese).

Deming, D. J. and Noray, K. (2020). Earnings dynamics, changing job skills, and STEM careers. *Quarterly Journal of Economics*, 135(4):1965–2005.

Duncan, G. J. and Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review*, 1(1):75–86.

Eckaus, R. S. (1964). Economic criteria for education and training. *Review of Economics and Statistics*, 46(2):181–190.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2024). GPTs are GPTs: Labor market impact potential of large language models. *Science*, 384(6702):1306–1308.

Felten, E. W., Raj, M., and Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 42(12):2195–2217.

Frank, R. H. (1978). Why women earn less: The theory and estimation of differential overqualification. *American Economic Review*, 68(3):360–373.

Frey, C. B. and Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114:254–280.

Giuntella, O., Lu, Y., and Wang, T. (2025). How do workers adjust to robots? evidence from China. *Economic Journal*, 135(666):637–652.

Goos, M. and Manning, A. (2007). Lousy and lovely jobs: The rising polarization of work in Britain. *Review of Economics and Statistics*, 89(1):118–133.

Green, C., Kler, P., and Leeves, G. (2007). Immigrant overeducation: Evidence from recent arrivals to Australia. *Economics of Education Review*, 26(4):420–432.

Hampole, M., Papanikolaou, D., Schmidt, L. D., and Seegmiller, B. (2025). Artificial intelligence and the labor market. NBER Working Paper 33509, National Bureau of Economic Research.

Hartog, J. (2000). Over-education and earnings: Where are we, where should we go? *Economics of Education Review*, 19(2):131–147.

Kuhn, P. and Shen, K. (2013). Gender discrimination in job ads: Evidence from China. *Quarterly Journal of Economics*, 128(1):287–336.

Li, J. (2016). Education-job mismatches and earnings among Chinese college graduates. *Chinese Journal of Sociology*, 36(3):64–85. (in Chinese).

Li, L., Wang, X., and Bao, Q. (2021). The employment effect of robots: Mechanisms and evidence from China. *Journal of Management World*, (9):104–119. (in Chinese).

Li, X. (2024). The impact of educational mismatch experience on firms' hiring decisions: Evidence from a resume audit experiment and online job postings. *Journal of Management World*, (4):121–145. (in Chinese).

Li, X., Lu, Y., and Wu, X. (2023). Educational mismatch among highly educated workers. *Educational Research*, (6):122–137. (in Chinese).

Liu, Y. (2019). The incidence of educational mismatch and skill mismatch and their income effects: An empirical analysis based on Chinese CGSS 2015. *Dongyue Tribune*, (3):60–68. (in Chinese).

Manuel, N. (2024). The use of major-related knowledge by early career college graduates. *Applied Economics*, 56(55):7302–7316.

Manuel, N. and Plesca, M. (2020). Skill transferability and the earnings of immigrants. *Canadian Journal of Economics*, 53(4):1404–1428.

McGuinness, S. (2006). Overeducation in the labour market. *Journal of Economic Surveys*, 20(3):387–418.

OECD (2017). *Getting Skills Right: Spain*. OECD Publishing, Paris.

Robst, J. (2007). Education and job match: The relatedness of college major and work. *Economics of Education Review*, 26(4):397–407.

Somers, M. A., Cabus, S. J., Groot, W., and Maassen van den Brink, H. (2019). Horizontal mismatch between employment and field of education: Evidence from a systematic literature review. *Journal of Economic Surveys*, 33(2):567–603.

Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. (2012). A new metrics for countries' fitness and products' complexity. *Scientific Reports*, 2(1):723.

Tsai, Y. (2010). Returns to overeducation: A longitudinal analysis of the US labor market. *Economics of Education Review*, 29(4):606–617.

Verdugo, R. R. and Verdugo, N. T. (1989). The impact of surplus schooling on earnings: Some additional findings. *Journal of Human Resources*, 24(4):629–643.

Verhaest, D. and Omey, E. (2006). The impact of overeducation and its measurement. *Social Indicators Research*, 77(3):419–448.

Wang, L., Qian, Y., Song, D., and Dong, Z. (2023). The job-transition effect of robot adoption and the characteristics of employment-sensitive groups: Micro-level evidence. *Economic Research Journal*, (7):69–85. (in Chinese).

Wang, Y. and Dong, W. (2020). How does the rise of robots affect China's labor market? evidence from listed manufacturing firms. *Economic Research Journal*, (10):159–175. (in Chinese).

Webb, M. (2020). The impact of artificial intelligence on the labor market. SSRN Working Paper 3482150.

Wolbers, M. H. J. (2003). Job mismatches and their labour-market effects among school-leavers in Europe. *European Sociological Review*, 19(3):249–266.

Xie, S., Wei, D., and Tang, Q. (2024). The impact of internet use on education–job matching: Evidence from CFPS 2016–2020. *Chinese Journal of Population Science*, 38(4):36–51. (in Chinese).

Yan, X., Zhu, B., and Ma, C. (2020). Industrial robot use and manufacturing employment: Evidence from China. *Statistical Research*, (1):74–87. (in Chinese).

Zhang, D., Yu, H., Li, L., Hu, J., Mo, Y., and Li, H. (2025). The measurement of AI exposure and its impact on labor demand in China: Evidence from large language models. *Journal of Management World*, 41(7):59–75. (in Chinese; English working paper version, 2026).

# Appendix

## A Related literature

The paper does three things: it reconstructs the exposure of the Chinese labor market to LLM technology; it documents vertical and horizontal mismatch in the Chinese labor market; and it examines how LLM technology affects that mismatch. This appendix reviews the literature on each of the three and situates the paper within it.

### A.1 Measuring occupational exposure to LLM technology

As artificial intelligence becomes embedded in economic life, a central difficulty for researchers and policymakers alike is how to measure its actual effects on industrial structure, the labor market, and job requirements in a way that is at once timely, precise, and systematic. Most existing work builds a quantifiable assessment framework around an *exposure index*, which converts the potential impact of AI on a given object—an occupation, an industry, a group of workers—into an interpretable number, providing a key handle for measuring the AI shock.

Within this literature, the task-based approach to constructing exposure measures is dominant. Grounded in the task-based model of work of Autor et al. (2003), it breaks with the tradition of treating an occupation as an indivisible whole: an occupation is first decomposed into a series of concrete, executable work tasks; the possibility and degree of substitution of human labor by AI is assessed at the task level; and task-level assessments are then aggregated upward under a consistent statistical logic into exposure indicators at the occupation, regional, or other levels of analysis.

Felten et al. (2021), Webb (2020), and Eloundou et al. (2024) combine task matching and text-semantic methods with expert or LLM scoring to construct occupation-level AI exposure indices, quantifying the substitution risk faced by different occupations, industries, and regions in the United States. They find that the impact of AI on the labor market is not confined to low-skill jobs; on the contrary, it tends to penetrate high-skill, cognition-intensive occupations (such as programmers and lawyers), and the substitution pressure is most pronounced in the middle-to-upper part of the wage distribution. Building on this framework, Zhang et al. (2025) adapt it to the Chinese context: drawing on large-scale recruitment data, they track the dynamics of the Chinese labor market from January 2018 to May 2024 and construct a task-based LLM exposure index. They find that as the new wave of AI centered on large language models accelerates, the growth of Chinese labor demand has slowed, while the aggregate LLM exposure of the labor market has declined steadily. Taken together, these studies place fine-grained task-level measurement at their core; they provide solid empirical support for analyzing the labor-substitution effects of AI and show that the new technology is reshaping the structure of the labor market through job requirements and skill matching. Relative to Zhang et al. (2025), this paper refines the construction of the index: the task-extraction step—previously a single step in which an LLM extracted standardized tasks directly from posting text—is decomposed into two steps, distilling specific tasks and then matching them to standardized tasks with the Sentence-BERT

framework. This markedly improves the interpretability and replicability of the construction, and hence the rigor and reliability of conclusions drawn from the index.

## A.2 Labor market mismatch

Educational mismatch refers to a discrepancy between a worker's education and the requirements of the job, and takes two main forms: vertical mismatch (an education level that is too high or too low, i.e., over- or under-education) and horizontal mismatch (a field of study that does not correspond to the field of work). A large literature shows that educational mismatch is widespread in labor markets around the world and carries significant economic consequences (Hartog, 2000). Duncan and Hoffman (1981) were the first to measure the wage effects of overeducation econometrically, finding a clear wage disadvantage for overeducated workers relative to workers whose education exactly matches their job; Verdugo and Verdugo (1989) subsequently confirmed this wage penalty. Meta-analyses across countries put the average wage loss from overeducation at roughly 10–20% (Hartog, 2000), while undereducated workers instead earn a certain wage premium for their scarce skills (Verdugo and Verdugo, 1989). Horizontal mismatch has received less attention, and its wage effects are smaller and less settled: the systematic review of Somers et al. (2019) finds a wage penalty for horizontal mismatch in general, but some studies find no significant wage reduction from working outside one's field. Overall, the earnings penalty from horizontal mismatch is generally weaker than that from vertical mismatch (McGuinness, 2006; Robst, 2007).

On measurement, the literature contains three main approaches to vertical mismatch. (1) Job analysis pre-assigns a required education level to each occupation in a classification and compares a worker's education against that standard; the approach dates to Eckaus (1964). (2) Realized matches computes the mean or mode of education among incumbents of an occupation from actual data and compares the individual against it—for example, education more than one standard deviation above the mean is classified as overeducation (Verdugo and Verdugo, 1989); this approach has been widely used in the subsequent literature on vertical mismatch. (3) Self-assessment asks workers directly what education their job requires, or whether they feel their education goes unused (Verhaest and Omey, 2006). Horizontal mismatch is measured in analogous ways, by subjective survey or objective matching. In the subjective approach, respondents themselves judge whether their field of study is related to their work (Bender and Heywood, 2011; Bender and Roche, 2013; Robst, 2007); the objective approach builds a field-to-job matching benchmark from standardized tools or classification frameworks, aiming to reduce the bias of subjective perception (Béduwé and Giret, 2011; Wolbers, 2003). Different measurement approaches produce different estimates of mismatch incidence. For rigor and transparency, this paper defines both vertical and horizontal mismatch by objective matching; the definitions are set out in detail in the main text.

In China, the massification of higher education has made educational mismatch more salient. Since the expansion of college enrollment in 1999, the number of college graduates has grown rapidly while high-skill jobs have grown more slowly, producing a structural oversupply of highly educated workers (Li et al., 2023). Recent estimates put the vertical mismatch rate in the Chinese labor market at about 24% and the field-of-study mismatch rate at about 34% (Li et al., 2023): close to one third of new graduates work in jobs

that either require less education than they have or lie outside their field of study. Nor does mismatch appear to be a short-lived transition: there is evidence that many workers face vertical or horizontal mismatch throughout their careers (Li et al., 2023; Wolbers, 2003). On consequences, a large literature at home and abroad focuses on the wage and career effects of educational mismatch. In general, relative to well-matched workers, both starting wages and subsequent wage growth are depressed for the overeducated and for those working outside their field (Li, 2016; Liu, 2019; Tsai, 2010). Liu (2019), using Chinese General Social Survey data, finds that overeducation significantly lowers wages while horizontal mismatch has no clear wage effect. Beyond wages, some studies point to a scarring effect on long-run career opportunities: the experimental study of Li (2024) finds that applicants whose résumés show a history of overeducation receive significantly fewer interview callbacks, suggesting that vertical mismatch lowers employers' confidence in an applicant's ability and creates a persistent disadvantage.

The causes of educational mismatch are complex, with both supply-side and demand-side components. On the supply side, educational attainment has risen quickly while the structure of jobs has adjusted slowly, leaving many graduates caught between jobs they cannot get and jobs they will not take (Li et al., 2023). Among individual characteristics, the young and labor market entrants are more prone to early-career vertical mismatch for lack of experience, and higher mismatch rates have also been observed for women and migrants (Frank, 1978; Green et al., 2007). On the demand side, some employers may hold biases or concerns about highly educated applicants—for example, that overqualified hires are more likely to quit—a possibility that has also been discussed in the Chinese context. Kuhn and Shen (2013), using data from a large Chinese job board, examine whether employers avoid hiring overqualified applicants. They find that bachelor's degree holders applying to postings requiring a junior college degree faced no significant discrimination, but when junior college graduates competed for jobs requiring secondary vocational or technical school education, being overqualified lowered their success rate. Vertical mismatch can thus harm employment outcomes in some circumstances, underscoring the structural difficulties Chinese college graduates face in the job market.

Overall, the existing literature amply documents the prevalence of educational mismatch and its negative effects on wages and employment opportunities. Several extensions remain worthwhile. First, most studies rely on labor force or graduate surveys and analyze mismatch rates and consequences at the aggregate level, without a detailed picture of the matching process between the two sides of the market. Using matched vacancy–applicant data from a large online recruitment market, this paper observes directly, at the micro level, how posting requirements line up with applicants' educational backgrounds, characterizing the overall state of educational mismatch in the current Chinese labor market and filling the gap on the supply–demand matching perspective. Second, the literature has paid little attention to how educational mismatch evolves under technological change. This paper brings that element in, using the matched data to analyze directly how an emerging technology shock alters matching efficiency in the labor market, offering a new perspective for the study of educational mismatch.

### A.3 Technological progress, AI, and labor market mismatch

How technological progress affects the labor market is a core question in economics. Classic studies show that the information technology revolution of the late twentieth century brought skill-biased technical change, raising the relative demand for high-skill labor and widening earnings gaps across skill groups (Acemoglu and Autor, 2011). The task-based model of technical change of Autor et al. (2003) further explains the shift in employment structure: computers and automation mainly displaced programmable, routine, middle-skill tasks, while leaving high-skill abstract cognitive tasks and low-skill non-routine manual service tasks hard to replace. The theory matches the job polarization observed in developed countries over recent decades: employment shares of middle-skill manufacturing and clerical jobs fell while shares of high-skill, high-paying occupations and low-skill service jobs rose together (Autor and Dorn, 2013; Goos and Manning, 2007). As automation deepened, workers in middle-income sectors were pushed toward both ends of the distribution, producing polarization (Autor and Dorn, 2013), with significant social consequences, including a shrinking middle-income workforce and rising labor market inequality (Autor, 2015). For China, Wang and Dong (2020), Li et al. (2021), Yan et al. (2020), Chen et al. (2022), and Giuntella et al. (2025) find that industrial robot adoption had a significant displacement effect on labor demand, reducing inflows of migrant labor into affected regions and depressing workers' wage income, with the negative shock concentrated among low-skill workers. Robot adoption also drove adjustments in job structure, shifting workers from strenuous to non-strenuous work and from routine to non-routine jobs (Wang et al., 2023).

AI differs from earlier waves of mechanization and computerization, which mainly displaced manual work and simple cognitive work. Large language models in particular possess sophisticated semantic understanding and cross-task learning, and can perform high-cognition tasks such as writing text, generating code, and analyzing information (Brynjolfsson and McAfee, 2014). The scope of substitution is thus extending from programmable work into highly specialized and creative, non-programmable work (Brynjolfsson and Mitchell, 2017; Zhang et al., 2025). Autor (2015) notes that although technological progress has historically created new jobs to offset those it destroyed, the rise of AI means that tasks traditionally performed by high-skill white-collar workers are beginning to be shared with automated tools, deepening anxiety about future employment; the empirical findings of Eloundou et al. (2024) support this judgment. Frey and Osborne (2017), estimating occupation-level automation probabilities, predicted that about 47% of existing jobs face a high risk of automation, sparking debate over mass AI-driven unemployment. The current consensus is that AI's employment effects are twofold (Autor, 2015; Acemoglu and Restrepo, 2019): on the one hand, AI substitutes for labor in some domains, reducing jobs or skill demand; on the other, it reshapes the task structure within jobs, creates new tasks and occupations (data annotators, algorithm-maintenance engineers), and raises employment indirectly through productivity gains (Bessen, 2019; Deming and Noray, 2020; Acemoglu et al., 2022). In the long run, the net effect on total employment and its structure depends on the balance of displacement and reinstatement (Acemoglu and Restrepo, 2019).

To date, however, academic research on how AI affects educational mismatch remains scarce. Existing studies mostly concern earlier waves of technology: for example, Xie et al. (2024) find that the spread of the internet reduces information asymmetry in job search and thereby the risk of vertical mismatch. The

literature on AI and the labor market approaches the question mainly from the demand side—the effects of AI on the quantity and structure of employment—with little attention to structural problems in the allocation of human resources on the supply side. Some scholars have begun to quantify AI's occupational impact and analyze its economic consequences: Felten et al. (2021) construct an occupational exposure index based on AI's competence at the tasks of different occupations, providing a tool for measuring the AI shock across industries and regions. In the Chinese context, Zhang et al. (2025) measure the LLM exposure of Chinese occupations and study its effect on labor demand, finding that high-exposure occupations exhibit demand upgrading: new postings in those occupations increasingly require high-skill, highly educated workers, while demand for low-skill positions contracts. AI progress is pushing the structure of labor demand toward higher skill. But demand is only half of the problem: whether workers on the supply side can upgrade their skills in time to meet job requirements equally determines the overall employment impact of technological progress. If education and training cannot keep pace with rising technical requirements, the gap between the speed of technical change and the speed of human capital adjustment shows up in the labor market as structural mismatch and wasted talent.

Against this background, the paper takes the labor supply side as its point of entry and studies how AI progress, represented by large language models, affects educational mismatch in the labor market and through what mechanisms. Bringing AI into the mismatch framework broadens the analysis of technology's labor market effects: previous discussion of AI has centered on employment quantities and skill structure, with little attention to matching efficiency and structural unemployment. The paper fills this gap, empirically identifying the effect of the LLM shock on educational mismatch rates. Its conclusions also enrich our understanding of AI's economic impact: AI progress not only changes the structure of job requirements but also reshapes allocative efficiency in the labor market through the application–recruitment matching process, offering a new perspective for a full assessment of AI's social and economic effects.

# B Construction of the LLM exposure index

The construction of the LLM exposure index in Zhang et al. (2025) has three stages: extracting standardized O*NET tasks from job postings, scoring each task's exposure, and aggregating the scores into an index weighted by market demand. This paper's methodological refinement, which combines the approach of Hampole et al. (2025) with the specific features of the posting data, breaks the task-extraction stage into four steps (Figure A1), making the whole procedure more transparent and traceable. The method keeps clean intermediate output at every stage of the processing pipeline, and it supports targeted refinement of the extraction results—for example, screening matches by a maximum-similarity threshold—thereby improving the stability and replicability of the results while preserving accuracy. The steps are as follows.

**Step 1: isolate task-relevant text.** Filter each posting down to the information relevant to work tasks. Posting text typically mixes work tasks, candidate requirements, benefits, and boilerplate courtesies; only the tasks and requirements bear directly on the index, and the redundant content would interfere with later processing, so it is discarded. This step uses a large language model with engineered prompts (Task 1 in

Section B.1 below). To keep the output logically consistent and stable, the prompt follows a five-part design: task description, detailed instructions, procedural guidance, output template, and reference examples.

**Step 2: distill specific tasks.** From the filtered text of Step 1, distill specific work tasks phrased in the semantic style of O*NET tasks, translated into English (Task 2 in Section B.1). Translation and rephrasing preserve the original meaning and detail of the Chinese posting as far as possible, to stay faithful to the job's true requirements.

**Step 3: clean and screen the tasks.** To remove invalid or low-quality task statements, screen the output of Step 2 further—for example, dropping statements too vague to carry information ("complete other tasks as required") and content unrelated to job duties. This step applies stricter screening and extraction rules (Task 3 in Section B.1) and ultimately filters out about 10% of the extracted task statements as invalid.

**Step 4: match to standard tasks.** Match the English work tasks surviving the first three steps to O*NET standard tasks by text similarity. Concretely, the Sentence-BERT framework encodes each posting task and each O*NET task as a 768-dimensional vector (using `paraphrase-MPNet-base-v2`, an open-source model that performs well on English matching), and cosine similarity is computed between the two. Each posting task is assigned the O*NET task with the highest similarity. The mean match similarity in this study is about 0.724, indicating high overall match quality.

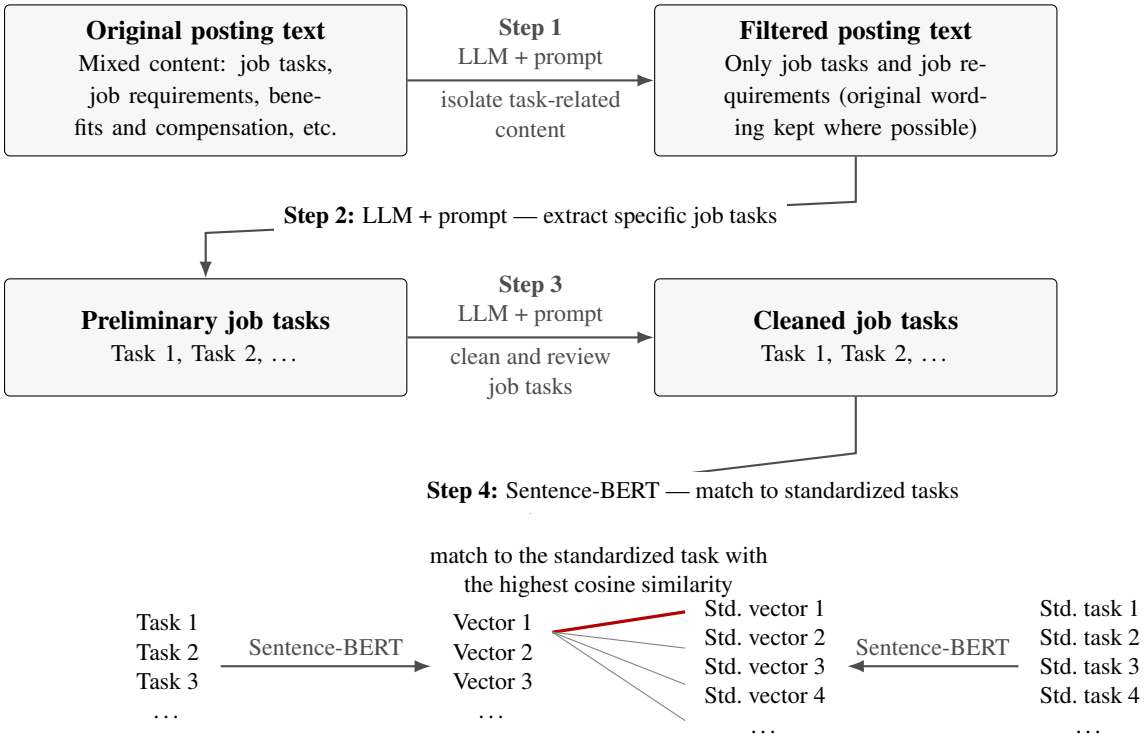



**Figure A1:** Steps for extracting standardized tasks from job postings

In practice, each posting that contains concrete job information yields on average 4.15 tasks. Building

on the task-level baseline exposure scores of Zhang et al. (2025) and the correspondence between SOC occupations and extracted tasks over 2018–2020, we aggregate to base-period exposure at the level of SOC broad occupations (6-digit), which serves as the basis for all subsequent analysis.[6] (Appendix C reports the exposure of the 50 SOC broad occupations that appear most frequently in the data.)

## B.1 Prompts used in task extraction

The prompts below are reproduced verbatim; following the convention of our earlier work, Chinese-language few-shot examples are replaced by an English gloss of what they demonstrate.

**Task 1: extract job tasks and requirements.**

```
[System] Your task is to review a job posting from an online recruitment platform and extract the job
tasks and requirements directly from the job posting while ensuring accuracy and structure. Please use
the following instructions, recommended steps, samples and output format to respond to user inputs.

[User] Instructions:
(1) The job posting may describe multiple aspects of the position, but you should focus only on job
    tasks (the activities and responsibilities the employee will be expected to perform) and
    requirements (the education, age, experience, qualifications, skills, or other personal
    characteristics the employer expects the candidate to have).
(2) Extract only the raw text corresponding to "job tasks" and "requirements." Ignore all other
    content, such as salary, benefits, company descriptions, and work arrangements.
(3) Classify correctly. Do not rely solely on section titles. Some sections labeled "Job
    Responsibilities" may actually describe required skills. Sometimes, job tasks or requirements may
    be scattered in separated places of the posting; make sure to capture all of the relevant
    information. Use the following rules: If a sentence describes actions performed on the job, it
    belongs in job tasks. If a sentence describes skills, qualifications, experience, age, gender, or
    other personal attributes, it belongs in the requirements.
(4) Pay attention to the negative semantics. Sometimes, in a recruitment advertisement, it states
    "There is no need to do a certain task." In such cases, do not categorize this statement under
    "job tasks" or "requirements".
(5) Ignore promotional or motivational content. Do not include statements about career paths, income
    potential, or generalized encouragement (e.g., "This could change your life!").
(6) Maintain original phrasing. Copy the relevant text verbatim. Do not summarize or paraphrase. If
    the information is scattered, group all relevant parts together under the appropriate section
    without modifying their structure.
(7) Handle ambiguous cases: At most times, if a sentence could belong to both sections, place it in
    the most relevant category based on primary intent. But some job tasks can be derived from a
    requirement statement. For example, from the requirement statement "Be proficient in using
    relevant architectural design software such as AUTO CAD, PHOTOSHOP, and 3D MAX", we know that the
    job includes the task "Use architectural design software such as AUTO CAD, PHOTOSHOP, and 3D MAX".
    In this case, classify this kind of statement as both a job task and a requirement.
(8) Avoid duplication. If the same information appears multiple times in different sections, include
    it only once in the most appropriate category.

A three-step process is suggested: First, filter out the original job posting and only keep
information about job tasks and job requirements. Second, reread the original job posting to ensure
```

[6] We use 2018–2020 as the base period for two reasons. First, the paper studies vertical and horizontal mismatch over 2021–2025, so the base period for the index should precede that window. Second, large language models had not yet diffused during 2018–2020, and the extraction results confirm that the occupational composition was relatively stable in that period, so exposure computed on it captures well an occupation's potential exposure to the LLM technology shock.

```
that no tasks or requirements are neglected in your answer. Third, check the job tasks and
requirements in your answer are correctly classified based on the instructions.

Output format (Do not include anything else in the answer. If no relevant information is found,
return "No relevant information found" under the respective section like "Job Tasks: No relevant
information found | Requirements: No relevant information found"): Job
Tasks:{1.Task1;2.Task2;...}| Requirements:{1.Requirements1;2.Requirements2;...}.

For reference, here is an example of correctly applied filters: <worked example on a Chinese-language
posting for a sales-administration assistant, showing four duties retained, six requirements
retained, and the benefits block discarded>

Here is the job posting you need to filter: <job post content>
```

### Task 2: format the work tasks.

```
[System] Your task is to extract specific tasks from the provided "Job Tasks" text in a recruitment
poster. Please use the following instructions, recommended steps, samples and output format to
respond to user inputs.

[User] Instructions:
(1) For structure and content, format each task to be similar to O*NET tasks, such as "Direct or
    coordinate an organization's financial or budget activities to fund operations, maximize
    investments, or increase efficiency", "Attend receptions, dinners, and conferences to meet
    people, exchange views and information, and develop working relationships".
(2) For length and language, each task that you extract is expected to be between 10 - 30 words in
    English.
(3) Maintain consistency with the original text description as much as possible.
(4) Some tasks in "Job Tasks" are long and complex such as "Operation, assembly, testing, inspection,
    packaging", split it into small tasks which are consistent with the O*NET tasks.
(5) Some tasks are simple and general such as "Product packaging", integrate and summarize moderately
    in combination with the context.
(6) Specific information such as place names, company names, and product names should not appear in
    the tasks you extract. Instead, generalized nouns and language should be used. For example, "Be
    responsible for the good credit of customers of Huabei and Jiebei of Ant Group" should be cleaned
    up as "Maintain good credit history of borrowers of the company's micro loan product". "1.
    Short-distance freight driver; 2. Fixed short-distance routes in Shenzhen, 2 to 3 trips per day,
    with each trip taking 4 to 5 hours" should be comprehensively summarized as "Drive short-distance
    freight along fixed routes".
(7) The aim is to calculate the text similarity between these extracted tasks and O*NET tasks.

A two-step process is suggested: First, extract specific tasks from the provided "Job Tasks" text and
organize the language and structure to make it comparable with O*NET tasks. Second, reread the given
"Job Tasks" text and check the extracted job tasks to ensure that there are no duplications or
omissions, and that the granularity and description of the tasks are consistent with those in O*NET.

Output format (Do not include anything else in the answer. There should be no quotation marks at the
very front and the very end of the output. If no relevant information is found, return "No relevant
information found"): {Task1 | Task2 | ...}.

For reference, here is an example of correctly applied filters: <worked example on a Chinese-language
posting for a course-sales role, yielding four task statements>

Here is the text from the recruitment poster: <job tasks and requirements>
```

**Task 3: re-clean the tasks.**

```
[System] Your task is to filter a "Job Tasks" list. Please use the following instructions, samples
and output format to respond to user inputs.

[User] Instructions:
(1) Some task descriptions contain non-English characters, please return "This task is
    multilingual."
(2) Some task descriptions are too broad, please return "This task is too broad." For example, the
    description "Complete other tasks assigned by the leader" is too general---we cannot determine
    the specific nature of the tasks from such a vague statement.
(3) Some task descriptions are actually requirements, please return "This task is a requirement."
    For example, the description "Master foundational JAVA and J2EE technologies" outlines a
    requirement rather than a specific, actionable task.
(4) Some task descriptions are overly detailed, including specific information like place names,
    company names, and product names. You should generalize such specifics and return the revised
    task. For example, "Be responsible for the good credit of customers of Huabei and Jiebei of Ant
    Group" should be generalized to "Maintain good credit history for borrowers of the company's
    microloan products." Similarly, "Travel to Japan for business assignments for one year" becomes
    "Travel to foreign countries for business assignments for one year."
(5) For qualified task descriptions, simply return the original text.

The input is a set of tasks like "Task1 | Task2 | ...". The output format should be (Do not include
the quotation mark): "Processed Task1 | Processed Task2 | ...".

For reference, here is an example of correctly applied filters: Original text: "<a Chinese-language
task statement: liaise with clients and negotiate cooperation> | Follow the leader's arrangements and
work conscientiously. | Have less than one year of experience. | Fixed short-distance routes in
Shenzhen, 2 to 3 trips per day, with each trip taking 4 to 5 hours | Make no more than 20 phone calls
daily to targeted clients." A good answer: "This task is multilingual. | This task is too broad. |
This task is a requirement. | Drive short-distance freight along fixed routes. | Make no more than 20
phone calls daily to targeted clients."

Here is the text you need to filter: <extracted tasks>
```

# C LLM exposure by SOC occupation

Table A1 reports the average LLM exposure of the 50 SOC occupations (6-digit) that appear most frequently on the recruitment platform over 2018–2020. Occupations are ordered by frequency, in descending order; exposure is the raw index, without standardization or other transformations.

**Table A1:** LLM exposure of the 50 most frequent SOC occupations

| SOC code | Occupation title | Exposure |
|---|---|---|
| 41-4012 | Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products | 0.657 |
| 11-2022 | Sales Managers | 0.723 |
| 43-4051 | Customer Service Representatives | 0.748 |
| 43-9061 | Office Clerks, General | 0.697 |
| 13-1071 | Human Resources Specialists | 0.654 |

*Continued on next page*

Table A1 (continued)

| SOC code | Occupation title | Exposure |
|---|---|---|
| 11-1021 | General and Operations Managers | 0.739 |
| 13-1161 | Market Research Analysts and Marketing Specialists | 0.810 |
| 11-2021 | Marketing Managers | 0.818 |
| 27-1024 | Graphic Designers | 0.798 |
| 43-6014 | Secretaries and Administrative Assistants, Except Legal, Medical, and Executive | 0.720 |
| 11-3121 | Human Resources Managers | 0.686 |
| 41-9041 | Telemarketers | 0.657 |
| 41-9022 | Real Estate Sales Agents | 0.574 |
| 13-2011 | Accountants and Auditors | 0.849 |
| 13-1081 | Logisticians | 0.614 |
| 11-3021 | Computer and Information Systems Managers | 0.846 |
| 11-3031 | Financial Managers | 0.834 |
| 17-2051 | Civil Engineers | 0.633 |
| 11-9021 | Construction Managers | 0.626 |
| 41-4011 | Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products | 0.672 |
| 13-1151 | Training and Development Specialists | 0.720 |
| 41-2031 | Retail Salespersons | 0.513 |
| 17-2071 | Electrical Engineers | 0.620 |
| 17-2141 | Mechanical Engineers | 0.580 |
| 43-4171 | Receptionists and Information Clerks | 0.534 |
| 27-3031 | Public Relations Specialists | 0.829 |
| 43-3031 | Bookkeeping, Accounting, and Auditing Clerks | 0.812 |
| 43-3071 | Tellers | 0.754 |
| 53-3032 | Heavy and Tractor-Trailer Truck Drivers | 0.274 |
| 11-2011 | Advertising and Promotions Managers | 0.857 |
| 11-3051 | Industrial Production Managers | 0.707 |
| 13-1051 | Cost Estimators | 0.807 |
| 11-9033 | Education Administrators, Postsecondary | 0.682 |
| 43-5081 | Stock Clerks and Order Fillers | 0.623 |
| 19-4099 | Life, Physical, and Social Science Technicians, All Other | 0.554 |
| 13-1023 | Purchasing Agents, Except Wholesale, Retail, and Farm Products | 0.725 |
| 41-1011 | First-Line Supervisors of Retail Sales Workers | 0.653 |
| 41-3031 | Securities, Commodities, and Financial Services Sales Agents | 0.665 |
| 11-3061 | Purchasing Managers | 0.797 |
| 11-9141 | Property, Real Estate, and Community Association Managers | 0.686 |
| 27-1025 | Interior Designers | 0.648 |
| 41-3021 | Insurance Sales Agents | 0.718 |
| 41-9031 | Sales Engineers | 0.692 |
| 53-7064 | Packers and Packagers, Hand | 0.353 |
| 33-9032 | Security Guards | 0.339 |
| 13-1199 | Business Operations Specialists, All Other | 0.800 |
| 13-1021 | Buyers and Purchasing Agents, Farm Products | 0.824 |
| 13-1111 | Management Analysts | 0.679 |
| 13-2052 | Personal Financial Advisors | 0.701 |

*Continued on next page*

Table A1 (continued)

| SOC code | Occupation title | Exposure |
|---|---|---|
| 25-2011 | Preschool Teachers, Except Special Education | 0.544 |

*Notes:* The table reports the average LLM exposure of the 50 SOC occupations (6-digit) appearing most frequently on the recruitment platform in 2018–2020. Occupations are ordered by frequency (descending); exposure is the raw index, without standardization or other processing.

# D The matched vacancy–applicant data

The data come from Zhaopin, founded in 1994 and one of the leading online recruitment platforms in China, with an online recruitment market share above 30%, about 374 million active individual users, and nearly 14.36 million firm users; this breadth and depth of coverage give the platform a degree of representativeness for the urban labor market.

The data cover January 1, 2021 to July 31, 2025, sampled from the demand side and matched to applicant information through application records. We first draw, for each year from 2021 to 2025, a random sample of 20,000 positions recruiting on the platform (the "posting sample"), with information including the posting date, job description, occupation, location, and industry. Building on the posting sample, we then extract all application records each posting received within its posting year (including the applicant and the application date), which constitute the "application sample." All job seekers appearing in the application sample constitute the "applicant sample"; for each applicant we extract basic characteristics—applicant ID, age, gender, education, and field of study—together with the most recent job held.

Each application also records an outcome: whether it reached a "near-hire" state. The platform cannot observe firms' hiring decisions directly, but it infers from the interaction records of the two parties whether firm and applicant have shown strong mutual interest—in the platform's internal terminology, a "positive reply." A positive reply is an intermediate step close to a completed hire: in the sample, the ratio of positive replies to the number of planned hires for the same posting averages 3.33, and for 51.5% of postings the number of positive replies is at or below the number of planned hires. In the analysis we therefore treat a positive reply as a near-hire state and use it as a proxy for a successful match between applicant and position.

We impose further restrictions on the analysis sample. First, the sampling design does not allow us to obtain application records from years other than the posting year, so postings published near the end of a year cannot be matched to their complete application history. We find that nearly 90% of applications arrive within two months of posting; the analysis therefore keeps only postings published in January–October of each year (January–May for 2025) and matches applications arriving within two months of posting. Second, we keep only full-time positions and their corresponding applications.

After these steps, the effective sample for the statistical analysis comprises 97,447 postings; 1,623,430 applications to those postings; and 1,439,854 distinct applicants behind those applications. Restricting applicants to those with junior college education or above leaves 1,452,440 applications from 1,279,223 distinct applicants. Panel A of Table A2 reports total applications and the positive-reply rate by year—the five-year average positive-reply rate is 19.44%—and Panel B reports the distributions of applicant gender,

age, and education by year.

**Table A2:** Descriptive statistics of the matched vacancy–applicant data

| | 2021 | 2022 | 2023 | 2024 | 2025 |
|---|---|---|---|---|---|
| *Panel A. Applications and replies* | | | | | |
| Total applications | 148,182 | 302,425 | 393,225 | 357,796 | 421,802 |
| Positive-reply rate (%) | 19.48 | 24.52 | 22.15 | 16.66 | 15.63 |
| *Panel B. Applicant characteristics* | | | | | |
| Share by gender | | | | | |
| Male (%) | 54.70 | 51.98 | 54.91 | 57.35 | 59.22 |
| Share by age | | | | | |
| 16–24 (%) | 21.39 | 29.29 | 25.30 | 22.95 | 19.19 |
| 25–34 (%) | 50.15 | 46.14 | 44.12 | 42.48 | 42.07 |
| 35–44 (%) | 22.05 | 18.95 | 23.20 | 25.61 | 27.87 |
| 45 and above (%) | 6.41 | 5.62 | 7.39 | 8.96 | 10.86 |
| Share by education | | | | | |
| Below junior college (%) | 9.16 | 9.92 | 9.27 | 8.81 | 7.03 |
| Junior college (%) | 32.97 | 34.01 | 32.46 | 31.80 | 30.50 |
| Bachelor's (%) | 51.50 | 50.67 | 53.08 | 53.62 | 55.58 |
| Master's/PhD (%) | 6.37 | 5.39 | 5.18 | 5.77 | 6.90 |

*Notes:* Panel A reports the total number of applications and the positive-reply rate in the matched vacancy–applicant data by year; Panel B reports the distribution of applicant gender, age, and education by year.

Note that of the 97,447 postings in the analysis sample, only 51.7% received at least one application (the "with-application sample"); postings without applications may be unattractive to job seekers or may receive applications through channels other than the online platform. A comparison shows that, relative to postings without applications, postings with applications have higher average LLM exposure (0.61 vs. 0.55, $p$-value $< 0.001$) and higher average salary (9627 vs. 9496, $p$-value $= 0.004$). Postings with applications have higher education requirements on average, concentrated at junior college and bachelor's degree, while the education requirements of postings without applications are concentrated in the no-requirement category. By occupation, postings with applications include more management and business-and-finance occupations, while postings without applications include significantly more production occupations. Because parts of the subsequent analysis are restricted to the with-application sample, we test before those analyses for the bias that sample selectivity could introduce (Appendix E compares the characteristics of postings with and without applications in detail).

# E Postings with and without applications

This appendix compares postings with and without applications. Table A3 reports summary statistics for LLM exposure and monthly salary in the two samples, and Table A4 reports $t$-tests and Mann–Whitney $U$-tests of the equality of exposure and salary across them. Figures A2–A5 show the differences between the two samples in the distributions of exposure, salary, required education, and SOC major group.

**Table A3:** Summary statistics of LLM exposure and monthly salary: postings with and without applications

| | LLM exposure | | Monthly salary (RMB) | |
|---|---|---|---|---|
| | Mean | Std. dev. | Mean | Std. dev. |
| Postings without applications | 0.5502 | 0.1915 | 9496.74 | 7118.52 |
| Postings with applications | 0.6115 | 0.1820 | 9627.66 | 7861.14 |

**Table A4:** Tests of equality of LLM exposure and monthly salary between postings with and without applications

| | $t$-test | | Mann–Whitney $U$ test | |
|---|---|---|---|---|
| | $t$ | $p$ | $U$ | $p$ |
| LLM exposure | −50.80 | 0.0000 | 939497517.00 | 0.0000 |
| Monthly salary | −2.87 | 0.0041 | 1190813714.50 | 0.0000 |

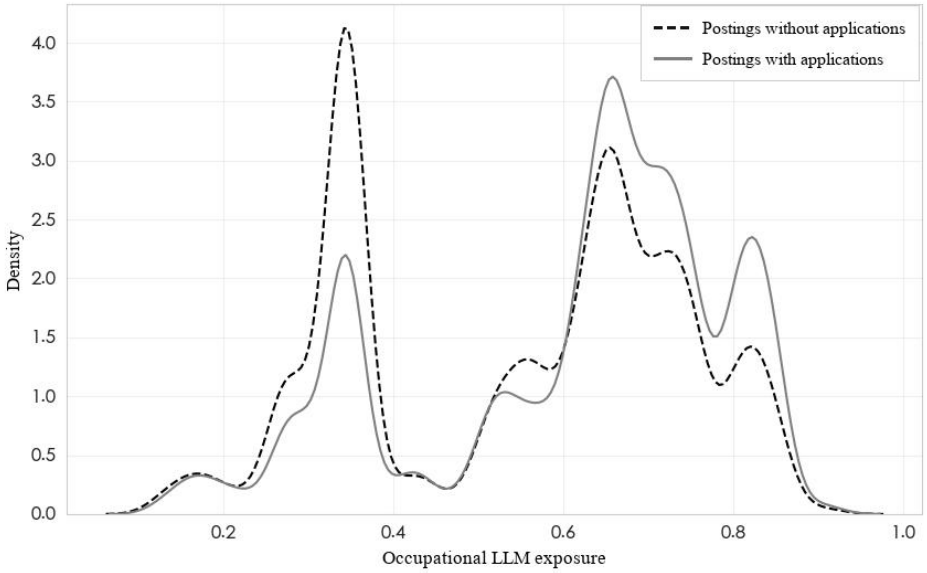



**Figure A2:** LLM exposure distributions of postings with and without applications

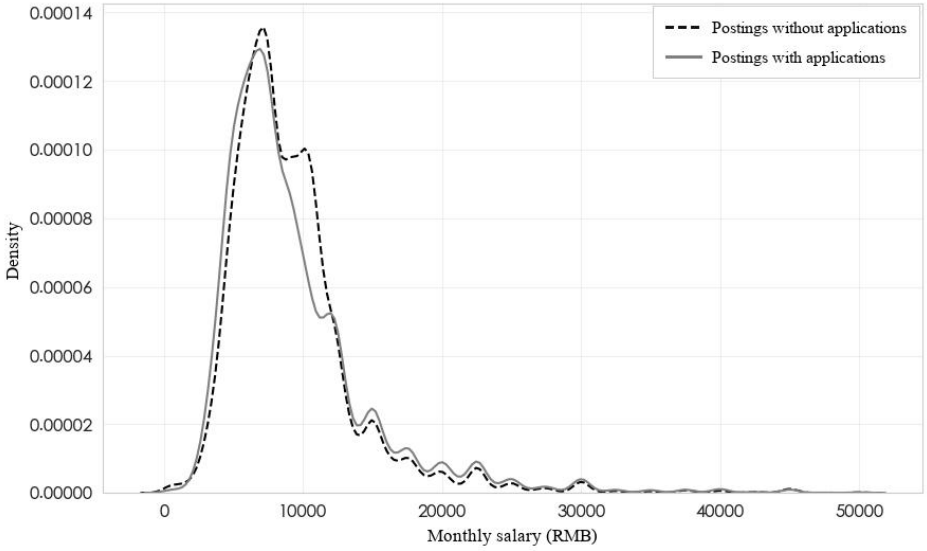



**Figure A3:** Salary distributions of postings with and without applications

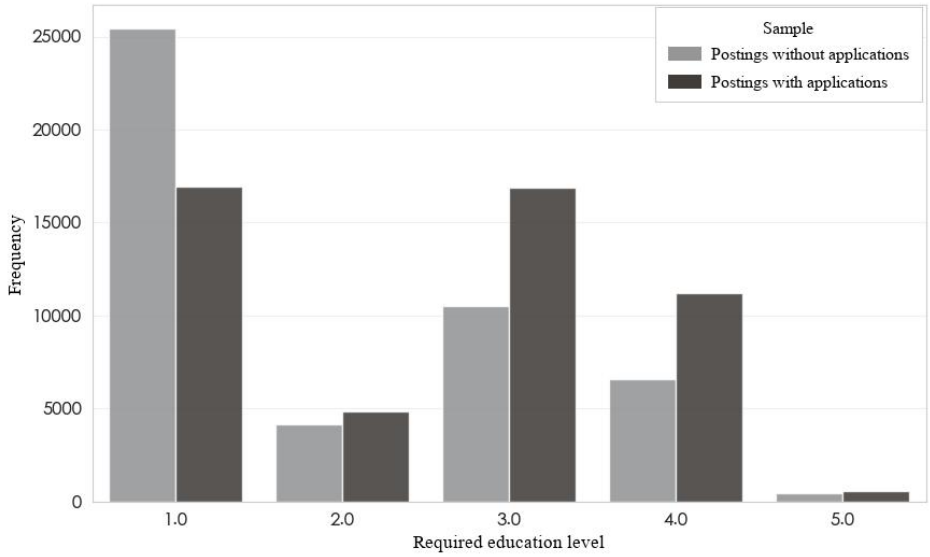



**Figure A4:** Required-education distributions of postings with and without applications

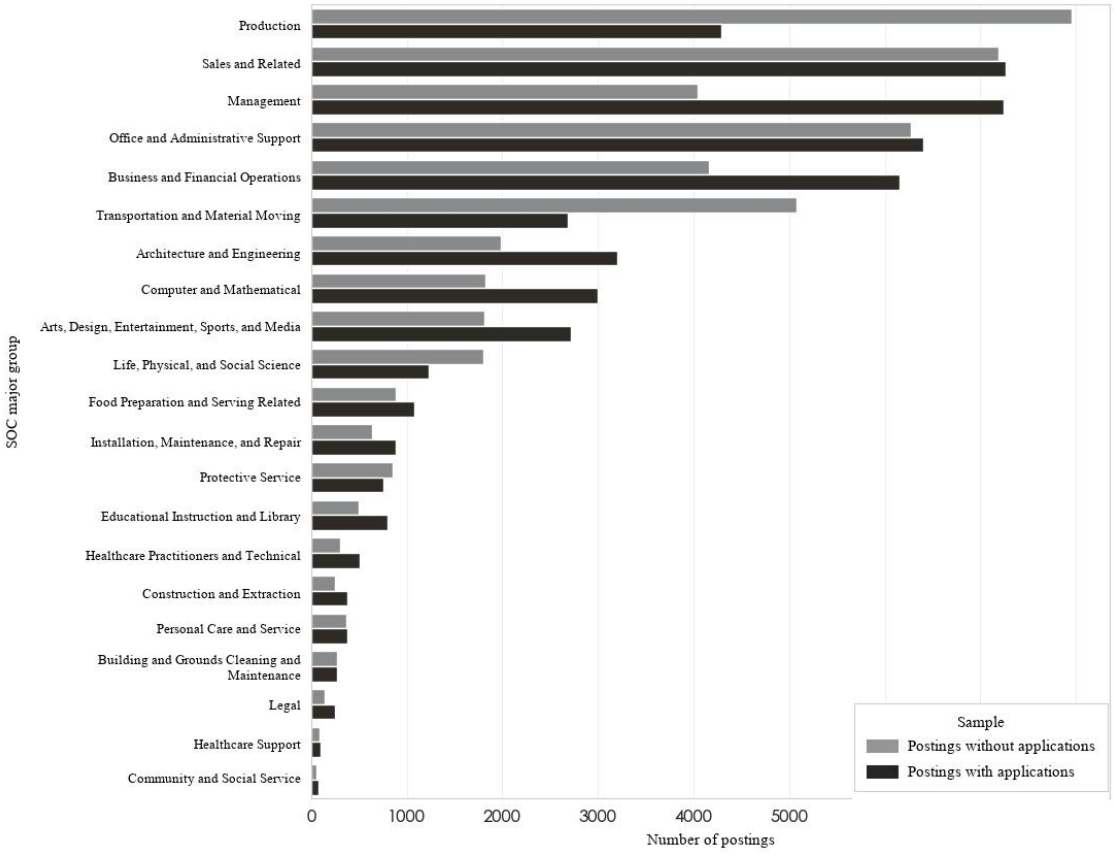



**Figure A5:** SOC major-group distributions of postings with and without applications

# F The SOC and CIP classification systems and the matching procedure

Constructing the mismatch measures first requires standardizing the occupations of the postings and the fields of study of the applicants. In line with established international practice, we map posting occupations to the SOC occupational classification and applicant fields of study to the CIP classification of instructional programs. This appendix describes both.

The Standard Occupational Classification (SOC) is the national standard for occupational classification in the United States, developed jointly by the Bureau of Labor Statistics (BLS), the Department of Labor's Employment and Training Administration (ETA), and other agencies to provide a unified framework for collecting, analyzing, and publishing occupational data, serving labor market statistics, employment services, and policymaking. It was first released in 1998; the latest version is the 2018 edition. The SOC comprises major groups (23, covering the broadest occupational domains, e.g., management occupations), minor groups (98, subdivisions of the major groups, e.g., "top executives" within management), broad occupations (459, groupings within minor groups, e.g., "chief executives" within top executives), and detailed occupations (867, specific occupation titles, e.g., "software developers"). Using LLM assistance combined with expert screening and review, we map the nearly one thousand occupation categories in the raw posting-sample data to SOC detailed occupations.

The Classification of Instructional Programs (CIP) is the US standard for classifying fields of study in higher education, developed by the National Center for Education Statistics (NCES) to unify the naming and classification of academic programs and to facilitate the collection, analysis, and comparison of education data. It was first released in 1980; the latest version is the 2020 edition. The CIP uses a four-level hierarchy with 6-digit codes (extended to 8 digits for some programs) to distinguish levels of detail: 2-digit major categories (the broadest fields, e.g., communication, journalism, and related programs), 4-digit subcategories (subdivisions of a major category, e.g., journalism and mass communication), 6-digit detailed programs (specific program names, e.g., journalism), and 8-digit program specializations (finer branches of some programs, e.g., journalism with a sports-journalism concentration). Using Sentence-BERT semantic-similarity matching combined with expert screening and review, we map the roughly forty-four thousand self-reported fields of study in the applicant sample to CIP detailed programs. Analyses involving fields of study are restricted to applicants with junior college education or above: their fields of study are more clearly defined, and they account for over 90% of the applicant sample. Moreover, since the CIP is designed to classify higher-education programs (associate through doctoral levels), it fits junior and senior high school and secondary vocational or technical credentials poorly.

# G Horizontal mismatch under alternative aggregation levels

Table A5 reports the average share of horizontal mismatch under alternative aggregation standards, crossing the CIP level used for fields of study (2-digit major category, 4-digit sub-category, 6-digit detailed program) with the SOC level used for occupations (major group, minor group, broad occupation). The shares in the table are computed without the imputation of missing labels.

**Table A5:** Average horizontal mismatch share under different aggregation standards

| | SOC major group | SOC minor group | SOC broad occupation |
|---|---|---|---|
| CIP major group (2-digit) | 28.9% | 40.2% | 56.0% |
| CIP submajor (4-digit) | 46.5% | 62.7% | 78.1% |
| CIP detailed program (6-digit) | 64.0% | 75.9% | 87.8% |

*Notes:* The shares in the table are computed without the missing-value imputation.

# H Details of the field-of-study standardization

A key step of the study is standardizing applicants' self-reported raw fields of study (44,461 after deduplication) into CIP programs. We use the Sentence-BERT framework to map each raw field name and each CIP standard program name into a 768-dimensional feature vector (because the task is bilingual, we use `paraphrase-multilingual-mpnet-base-v2`, the open-source model that currently performs best on multilingual text); each raw field name is then matched to the CIP program whose feature vector has the highest cosine similarity with it. Figure A6 shows the distribution of the maximum similarity attained by each raw field. Low-similarity matches—which may arise from irregular phrasing in applicants' self-reports, or because the CIP system contains no standard program that corresponds well to the raw field—should not, in principle, be treated as valid matches and need to be removed. The distribution in Figure A6 shows a clear kink near a similarity of 0.5, with a flat long tail to its left, so we set the threshold at 0.5. In addition, matches with similarity above 0.5 are further screened through expert review.

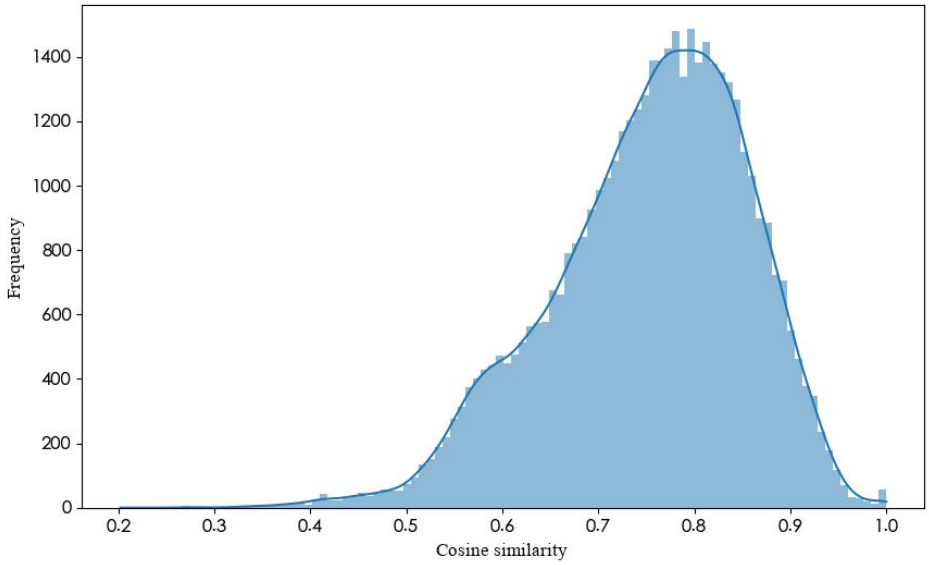



**Figure A6:** Distribution of semantic similarity in the field-of-study standardization

# I Imputing missing mismatch labels

This appendix describes how missing values of the "downward application" and "cross-field application" labels are predicted.

We use the applications that carry downward-application and cross-field-application labels as training samples and apply machine learning to predict the labels for applications where they are missing. In choosing the model, given the large feature space and a degree of class imbalance in the labeled samples, we settled on a support vector machine (SVM). The algorithm generalizes well in high-dimensional, small-sample classification settings and effectively guards against overfitting. We construct a 137-dimensional feature vector from the posting and applicant characteristics of each application record, applying one-hot encoding to categorical features (such as industry and field-of-study major category) and $z$-score standardization to numerical features (such as salary and years of work experience) to remove scale effects.

Table A6 reports performance on the test set. Prediction accuracy is 0.79 for downward application and 0.69 for cross-field application.

**Table A6:** Prediction performance on the test set

| Class / metric | Precision | Recall | F1-score |
|---|---|---|---|
| *Prediction of "downward application"* | | | |
| Class-0 sample | 0.79 | 0.87 | 0.83 |
| Class-1 sample | 0.79 | 0.68 | 0.73 |
| Macro average | 0.79 | 0.78 | 0.78 |
| Weighted average | 0.79 | 0.79 | 0.79 |
| *Prediction of "cross-field application"* | | | |
| Class-0 sample | 0.70 | 0.83 | 0.76 |
| Class-1 sample | 0.65 | 0.48 | 0.55 |
| Macro average | 0.68 | 0.65 | 0.65 |
| Weighted average | 0.68 | 0.69 | 0.67 |

*Notes:* Prediction accuracy on the test set is 0.79 for "downward application" and 0.69 for "cross-field application."

# J Heterogeneity in labor market mismatch

We examine how vertical and horizontal mismatch are distributed across occupations. Figure A7 shows, for each level of required education, the five occupations with the highest and the five with the lowest shares of vertical and of horizontal mismatch.[7] The results reveal sharp heterogeneity in both forms of mismatch across occupations.

Professional-service positions—law, education, and science-related jobs—have the highest vertical mismatch rates: the corresponding junior college positions are prone to being crowded by bachelor's degree applicants, and the corresponding bachelor's positions likewise attract applicants with still higher degrees.

[7] Occupations are aggregated to SOC minor groups; for rigor, occupations with fewer than 100 observations are excluded from the comparison.

But the very specialization of these occupations keeps their horizontal mismatch low; whether at the junior college or the bachelor's level, they are rarely contested by applicants from other fields.

Blue-collar jobs among junior college positions do not attract bachelor's degree applicants and show little vertical mismatch; but because their skill content is less specialized, they also exhibit the most severe horizontal mismatch within the pool of junior college applicants. Similarly, among bachelor's positions, production and manufacturing occupations are the ones with the most severe horizontal mismatch.

Management and clerical occupations among bachelor's positions hold little attraction for applicants with advanced degrees and likewise maintain low vertical mismatch, being filled essentially by bachelor's degree applicants.

We further examine heterogeneity in vertical and horizontal mismatch by applicant age and education. Applicants are grouped into five age brackets: 16–24, 25–29, 30–34, 35–44, and 45 and above. On education, vertical mismatch is shown for four levels—secondary vocational/technical school or high school, junior college, bachelor's degree, and master's/doctoral degree—and horizontal mismatch for three levels: junior college, bachelor's degree, and master's/doctoral degree. Figure A8 shows the average vertical and horizontal mismatch shares across applicant groups, with darker shades indicating more severe mismatch.

The vertical mismatch pattern in Panel A of Figure A8 is nonlinear in education—low in the middle, high at both ends. Junior college applicants have the lowest overall vertical mismatch, most of them matching to junior college positions; applicants with higher or lower education experience more severe mismatch, with high school applicants flowing heavily into positions with no education requirement and applicants with bachelor's degrees or above flowing broadly into junior college positions. By age, the average mismatch shares of the 16–24 youth group and the 45-and-above group are significantly higher than those of the other age brackets, reflecting the disadvantaged position of the youngest and oldest job seekers in the labor market.

The horizontal mismatch pattern in Panel B of Figure A8 looks different. In contrast to vertical mismatch, junior college applicants have the highest horizontal mismatch shares, possibly because their vocational skills are relatively basic and their occupational choices more flexible. Across age, the horizontal mismatch share of junior college applicants is essentially stable, that of bachelor's degree applicants rises slightly with age, and that of master's and doctoral applicants declines markedly with age.

# K Prompts used in skill extraction

This appendix reproduces the prompt used in the skill-extraction step. As in Appendix B, Chinese-language few-shot examples are replaced by an English gloss.

**Task 1: extract skill requirements.**

```
[System] Your task is to extract specific skills from the provided "Job Tasks" and "Requirements"
text in a recruitment poster. Please use the following instructions, recommended steps, samples and
output format to respond to user inputs.

[User] Instructions:
(1) Definition of Skill: Skills are people's abilities, knowledge, and expertise that are needed to
```

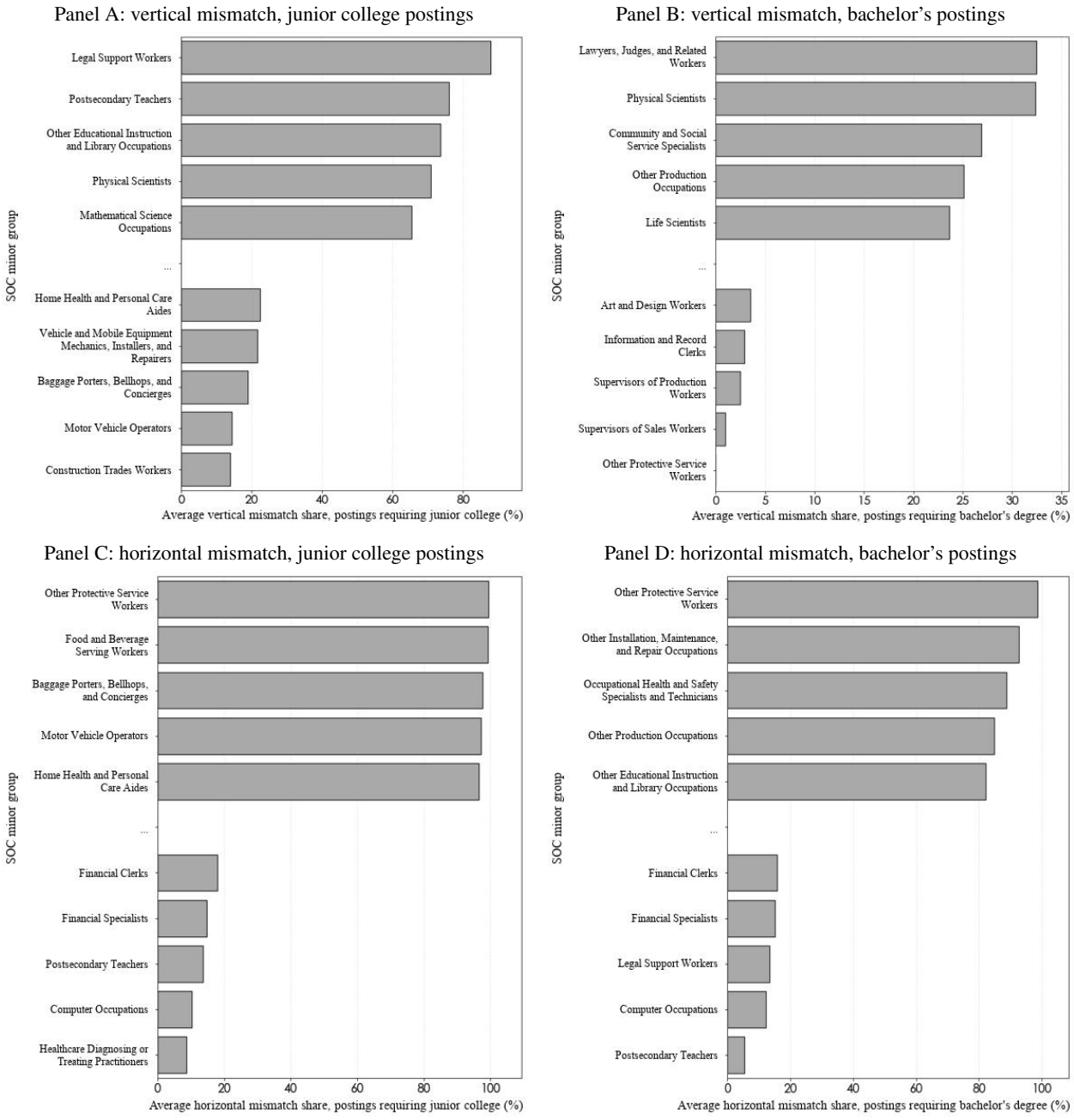
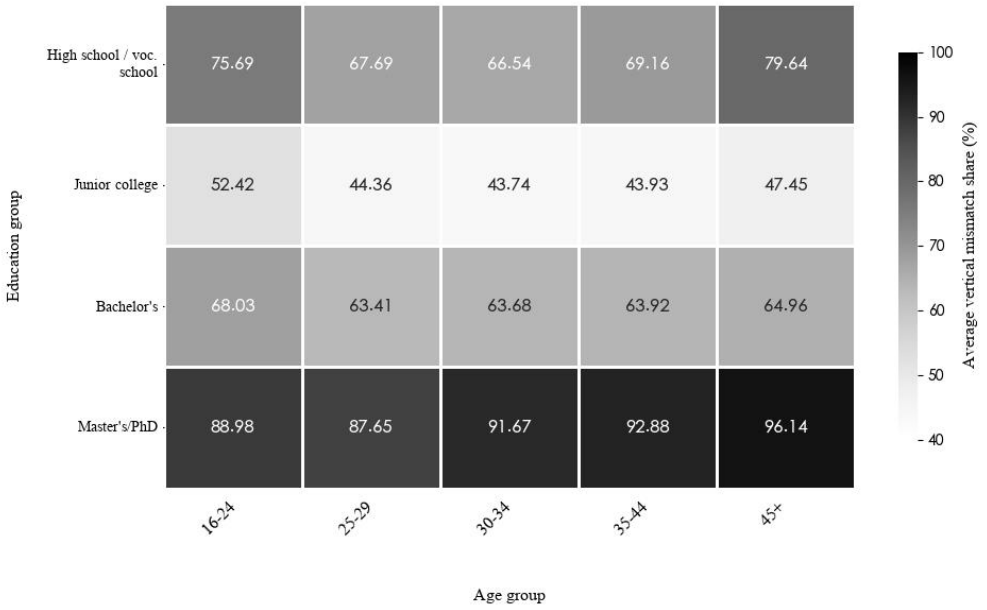



**Figure A7:** SOC minor groups with the five highest and five lowest average mismatch shares, by required education

Panel A: vertical mismatch




Panel B: horizontal mismatch

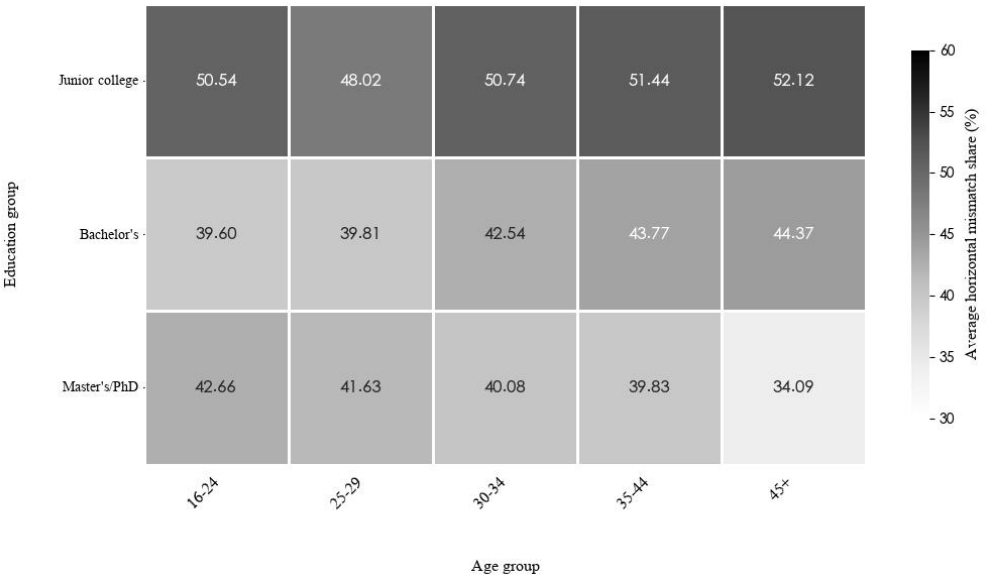



**Figure A8:** Average mismatch shares by applicant age and education

```
    carry out specific tasks or activities effectively. People can acquire these skills through
    formal education, training, and experiences gained in different settings such as work, volunteer
    opportunities, and personal interests.
(2) According to lightcast standard, there are mainly three types of skills: (a) Common Skills are
    prevalent across many different occupations and industries, including both personal attributes
    and learned skills. (e.g. "Communication" or "Microsoft Excel"). These include soft skills, human
    skills, and general competencies. (b) Specialized Skills are more specific. They're primarily
    required within a subset of occupations or equip one to perform a specific task (like "NumPy" or
    "Hotel Management"). These are sometimes known as technical skills or hard skills.
    (c) Certifications are recognizable qualification standards assigned by industry or education
    bodies (including "Cosmetology License" or "Certified Cytotechnologist").
(3) Notes for extraction:
    - Identify explicitly mentioned skill terms and terminology from Requirements. Some requirements
      have the form like "Able to do ..., capable of doing ..." This may be a skill statement.
    - Identify implied skill requirements within task statements. Some skills can be inferred from
      job tasks. For example, from expression "Responsible for requisition, distribution, and
      inventory management of daily office supplies", we can know the job need skills of inventory
      management and record keeping.
(4) Break down compound skills into basic skill units. In statement "Able to express ideas clearly
    and communicate effectively, capable of analyzing and solving problems efficiently", there are
    actually two skills "ideas expression and communication" and "problems solving".
(5) The original text is Chinese, but the extracted skills should be English.
(6) From each text, extract the main skills---at least one, but no more than 15 items. The formation
    of each skill should be like skills expression in lightcast (e.g., video remote interpreting,
    pricing feedback, embedded value accounting). If there are no skills, return "no relevant
    information found" --- but avoid using this response lightly; try to infer any possible skills
    before concluding.
(7) The extracted skills should align as closely as possible with the original meaning.

A two-step process is suggested: First, extract skills from the provided text based on the
instructions and organize the language and structure to make it comparable with lightcast skills.
Second, reread the given text and check the extracted skills to ensure that there are no duplication
or omissions.

Output format (Do not include anything else in the answer. There should be no brace at the very front
and the very end of the output. If no relevant information is found, return "No relevant information
found"): {Explicit Skills: Skill1 | Skill2 | ...; Implied Skills: Skill1 | Skill2 | ...}.

For reference, here is an example of correctly applied extraction: Original text: <worked example on
a Chinese-language accounting-clerk posting with five duties and eight requirements> A good answer:
"Explicit Skills: Accounting | Accounts Payable | Accounts Receivable | Analytical Skills |
Communication | Ethical Standards And Conduct | Financial Regulations | Reconciliation | Software
Systems | Tax Preparation | Willingness To Learn; Implied Skills: Management | Settlement".

Here is another example: Original text: <worked example: a posting with no task information whose
requirements ask for familiarity with real-estate cashier work or a background as a retired bank
employee> A good answer: "Explicit Skills: No relevant information found; Implied Skills: Operations
| Real Estate".

Here is the text from the recruitment poster: <job tasks and requirements>
```