Now Hiring: Firm-Level Adoption of Generative AI in China*

Lixing Li[1], Zhencen Liu[2], Anqi Wang[1], He Wang[2], and Shu Yu[†2]

[1]National School of Development, Peking University

[2]World Bank

June 4, 2026

Abstract: This paper uses job posting data to examine the adoption of generative artificial intelligence (GenAI) among Chinese firms. Drawing on 72 million job advertisements posted on China's online recruitment platforms, we track the hiring activities of over 5,000 listed firms and approximately 2.5 million non-listed firms from 2015 to 2025. GenAI adoption is measured with a keyword-based classification framework that groups relevant skills into three hierarchical categories: knowledge about GenAI applications, technical terminology specific to GenAI, and AI-related skills predating GenAI. Early evidence points to a rapid yet uneven diffusion pattern among firms in China. First, GenAI adoption remains higher among larger and more technically sophisticated firms: 21 percent of listed firms and 0.2 percent of non-listed firms had recruited GenAI-related roles by the end of 2025, with adoption accelerating after the release of DeepSeek-R1. Second, firm size, market service sectors, and prior adoption of earlier AI technologies are each positively associated with both the propensity and intensity of GenAI adoption. Third, adoption sophistication is highest among listed firms in market services, and a comparison with the United States shows that the cross-country adoption gap is much larger among non-listed firms than among listed firms.

JEL codes: O33, O57, J23, D22, C81

Key words: Generative AI, Measuring AI Adoption, Technology Diffusion, Chinese Firms

*Acknowledgments: The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. We thank Aart Kraay, Gaurav Nayyar, Steven Pennings, Elitza Mileva, Yusha Li, Sharif Kazemi, Samuel Fraiberger, and other seminar participants at the World Development Report, the Prosperity Vertical Chief Economist seminar at the World Bank, and National School of Development at Peking University for helpful comments and suggestions. We are grateful to Xiaobo Zhang, Zheng Song, and Xiaohan Yang for sharing their data, and to Jiayu Gao for excellent research assistance. This research received support from the World Bank Research Support Budget Fund.

[†]Emails: lilixing@nsd.pku.edu.cn (Lixing Li; corresponding author); zliu12@worldbank.org (Zhencen Liu); anqiwang@stu.pku.edu.cn (Anqi Wang); hwang21@worldbank.org (He Wang); syu2@worldbank.org (Shu Yu).

1 Introduction

China has emerged as one of the world's leading AI powers. Yet systematic firm-level evidence on the speed and depth of AI adoption among its firms remains scarce. Existing studies draw predominantly on data from advanced economies, such as the United States and Europe (Acemoglu et al. 2022; Babina et al. 2023), where firm-level technology sophistication and corporate governance structures differ markedly from those in emerging markets. In China, distinct market features, including a large manufacturing base, a prominent state presence, and rapidly expanding digital infrastructure, may shape both the capacity and the incentives for AI adoption in ways that differ from those in advanced economies. Given China's growing role in the AI value chain, filling this gap is vital to understanding how AI diffuses across firms in a developing economy.

While previous studies focused on using exposure-based measures to identify occupations or tasks that could be affected by (Gen)AI in China (Chen et al. 2025; Zhang et al. 2025), we focus on firm-level adoption, where firms incorporate GenAI-related capabilities into their operations through hiring. Following previous research,[1] a firm is defined as an GenAI adopter after it first posts a vacancy online explicitly requiring the use of GenAI applications or technical skills. Here, we combine job postings from three major Chinese online job boards and match them with approximately 5,000 listed firms, while vacancies on 51job.com are matched to about 2.5 million non-listed firms. A systematic keyword-based approach is applied to identify postings that require GenAI-related skills. To ensure identification quality, the approach addresses cross-platform duplicates, tailors keyword selection to the Chinese context, mitigates false positives, and classifies postings by business function to capture the sophistication of adoption. It was applied to 72 million job postings from 2015 to 2025 to provide the first comprehensive job-posting-based picture of GenAI adoption patterns among Chinese firms.

Measuring GenAI adoption is inherently challenging as a firm can engage with the technology across multiple dimensions, including invention, signaling, perception, and actual use (Babina 2026; Cirera, Comin, and Cruz 2026). Currently, there are four main estimation approaches in the literature. Survey-based measures ask firms or workers directly and yield interpretable adoption rates, but estimates swing widely across studies, depending on the technologies covered, question

[1] See, for instance, Babina et al. (2024), Hampole et al. (2025), and Hosseini and Lichtinger (2025).

wording, sampling frame, and respondents.[2] Patent-based measures identify AI invention and tend to capture technology developers and suppliers rather than ordinary adopters (Calvino et al. 2022). Self-disclosure measures, built from AI-related keywords in annual reports or filings, offer broad coverage and are common in the Chinese setting, but they reflect strategic orientation and signaling as much as operational use (Yao et al. 2024). Finally, there is a growing job-posting literature that infers adoption from firms' demand for AI-related skills.[3] As recruiting AI talent is a costly and observable organizational action, this approach captures firms' concrete steps toward AI integration at high frequency and with detail on the business functions where it is deployed. The approach provides valuable insights into firm-level adoption patterns, including potential new job creation. Therefore, despite its limitations in capturing adoption through off-the-shelf tools or retraining existing staff, we adopt this job-posting-based approach and apply it at scale in a developing-country setting to both listed and non-listed firms.

The evidence from the job-posting approach suggests a rapid yet highly uneven pattern of diffusion in China. First, GenAI adoption among Chinese listed firms rose steadily following the release of ChatGPT, reaching approximately 11 percent by the end of 2024 and 21 percent a year later. Adoption accelerated sharply after the release of DeepSeek in early 2025,[4] particularly among non-listed firms, consistent with the fact that DeepSeek is a domestic, low-cost, and readily localized model that lowered adoption barriers for smaller firms. Nevertheless, adoption remains substantially higher among listed firms (21 percent) than non-listed firms (0.2 percent) by the end of 2025. Second, firm size, market service sectors, and prior adoption of earlier AI technologies are each positively associated with both the propensity and intensity of GenAI adoption. Third, adoption sophistication is highest among listed firms in market services. Finally, a comparison with the United States using the same job-posting-based approach shows that GenAI adoption rates among Chinese listed firms are broadly comparable to those of their U.S. counterparts, while adoption among non-listed firms remains substantially lower in China.

Our paper contributes to the existing literature in several ways. It is the first to document GenAI adoption patterns among firms in a developing country with detailed granularity. Some studies used a survey-based approach to measure individual adoption of GenAI in advanced

[2] See, for instance, McElheran et al. (2024), Bick, Blandin, and Deming (2025), and Yotzov et al. (2026).
[3] See, for instance, Babina et al. (2024), Hampole et al. (2025), Alekseeva et al. (2025), and Shi et al. (2025).
[4] Specifically, we refer to the release of DeepSeek-R1 on January 20, 2025.

economies (such as Bick, Blandin, and Deming 2025). Others, such as the World Bank Enterprise Surveys, have used firm-level surveys to track firms' adoption of GenAI in developing countries by interviewing managers who may not have a clear view of how GenAI has been integrated across business functions (World Bank forthcoming). Our paper also contributes to the general literature on technology adoption (e.g., Comin and Hobijn 2004; Cirera, Comin, and Cruz 2026).

Second, this is the first rigorous study that tailors the job-posting approach to a developing-country context. Existing job-posting measures of AI adoption have been developed primarily for the United States and other advanced economies, where postings are often collected from relatively standardized sources and linked to firms through well-established identifiers. Applying this approach to China requires addressing a different set of measurement challenges. We combine postings from three major online job boards, remove cross-platform duplicates, match vacancies to both listed and non-listed firms, construct GenAI keyword lists that include domestic tools and Chinese-language terminology, and manually validate the resulting classifications. This procedure allows us to adapt a widely used measurement strategy to a setting where firm identifiers, posting formats, and language conventions are less standardized, thereby extending the empirical toolkit for studying technology adoption in developing economies.

Third, it is one of the first studies that attempts to document new GenAI-related job titles and occupations appearing in online vacancies in a developing country. Since job postings include detailed information about job titles, responsibilities, and required skills, they enable us to go beyond simply measuring whether firms adopt GenAI and to analyze how such adoption affects the composition of labor demand. We distinguish postings that require GenAI technical skills from those that require GenAI tool knowledge, and use this distinction to document emerging patterns in GenAI-related labor demand. Technical-skill postings are concentrated among ICT specialists, especially algorithm engineers and software developers, while tool-based postings extend to non-ICT occupations such as operations specialists, UI designers, graphic designers, and administrative clerks. This evidence shows that GenAI-related hiring is not limited to technical functions but extends to broader business activities, providing early evidence on how the technology is associated with changing occupational demand in a developing-country labor market.

The remainder of the paper is organized as follows. Section 2 reviews the existing approaches used to measure AI and GenAI adoption and clarifies how job-posting measures differ from

survey-, patent-, and disclosure-based approaches. Section 3 describes datasets and procedures used to identify GenAI-related vacancies, including deduplication, keyword construction, validation, and business-function classification. Section 4 presents stylized facts on GenAI adoption among Chinese firms, examines how adoption varies by firm characteristics, and documents emerging patterns in GenAI-related labor demand. Section 5 compares the Chinese patterns with those observed in the United States using the same job-posting-based approach. Section 6 concludes and discusses the limitations of current evidence.

2 A Review of Existing AI Adoption Measures

Measuring GenAI adoption is conceptually challenging as available data capture different dimensions of firms' engagement with the technology.[5] The existing literature focuses on four estimation approaches, each with its own limitations and advantages. Survey-based measures can directly estimate the fraction of firms or workers using AI, but the resulting estimates are sensitive to the definition of AI, the respondent, the sampling frame, and the weighting scheme. Patent- and disclosure-based measures do not usually provide clear adoption rates, as the former captures innovation and the latter indicates strategic moves or market signals. In contrast, job-posting measures capture a concrete organizational action, the costly recruitment of workers with AI-related capabilities, which is more closely tied to firms' efforts to integrate AI into their organizational functions.[6] Table 1 lists a brief comparison across the four approaches.

[5] As emphasized by Babina (2026), different AI datasets capture distinct economic objects, including invention versus use, internal capability building versus outsourced adoption, realized activity versus external perceptions, and broad organizational deployment versus narrowly defined applications (Cirera, Comin, and Cruz 2026).

[6] Notably, a small but particularly informative strand of research observes firms' actual activation or use of specific AI tools through administrative data. For example, Yuan (2025) studies the staggered adoption of an AI tool introduced by a large Chinese e-commerce platform in 2024. Administrative usage data provide an important benchmark for actual adoption, but they typically concern one commercial ecosystem or a narrow class of tasks and are rarely available at the economy-wide scale.

Table 1: Measurement Approaches to AI and GenAI Adoption

Measurement approach	Representative studies	What the measure mainly captures	Main advantage and limitation
Survey-based	McElheran et al. (2024); McElheran et al. (2025); Bick, Blandin, and Deming (2025); Bonney et al. (2026); Yotzov et al. (2026)	Self-reported AI use by firms, establishments, executives, or individual workers	Provides direct and interpretable evidence on reported use. Estimates are sensitive to the technology covered, question wording, sampling frame, weighting scheme, and respondent type. Worker surveys additionally capture individual use rather than formal firm adoption.
Patent-based	Cheng, Lee, and Tambe (2023); Calvino et al. (2022)	AI invention, technological development, or commercialization by firms	Provides an objective signal of innovative activity. Identifies AI developers or innovators rather than ordinary firms adopting externally supplied AI tools.
Self-disclosure-based	Calvino et al. (2022); Yao et al. (2024)	AI-related strategic orientation, public communication, investment intentions, or market signaling	Offers broad coverage and can be linked to firm outcomes over time. AI-related text may contain strategic rhetoric or forward-looking statements.
Job-posting-based	Calvino et al. (2022); Alekseeva et al. (2025); Shi et al. (2025); Hosseini and Lichtinger (2025)	Firms' observable investment in AI-related human capital and efforts to integrate AI into organizational functions	Captures function-specific organizational action at relatively high frequency. May miss adoption through off-the-shelf tools, external providers, or existing employees; vacancies may reflect intended rather than realized implementation.

2.1 Survey-based approach

The most direct approach to measuring AI adoption is to ask firms (or workers) whether they use AI in their operations. Survey-based measures yield an interpretable extensive-margin indicator—the share of firms currently using AI—and, when based on representative samples, allow researchers to characterize adoption across firm size, industry, geography, and business functions. However, reported adoption rates vary substantially across surveys as they differ in sample composition and weighting, question wording and technological scope, and respondent incentives

and information sets.

Evidence from the United States illustrates this sensitivity (Crane, Green, and Soto 2025). Using the 2018 Annual Business Survey, McElheran et al. (2024) find that fewer than 6 percent of firms used at least one of the five early AI technologies in 2017, though the employment-weighted rate exceeded 18 percent. Among manufacturing firms, McElheran et al. (2025) report adoption ranging from 13 to 23 percent in 2021 depending on the definition used. After the diffusion of GenAI applications, the Census Bureau's Business Trends and Outlook Survey reports that 18 percent of U.S. firms used AI during November 2025–January 2026, rising to 32 percent when weighted by employment (Bonney et al. 2026). At the other extreme, a survey of senior executives using a broad definition of AI reports a U.S. firm adoption rate of 78 percent (Yotzov et al. 2026).[7]

International evidence displays similar variation. Based on harmonized enterprise surveys, Eurostat reports that 20 percent of EU enterprises used at least one AI technology in 2025, up from 8 percent in 2023, with adoption rates ranging from 5 percent to 42 percent across member states (Eurostat 2025). Using executive surveys with comparable questions across four countries, Yotzov et al. (2026) report much higher adoption rates: 78 percent in the U.S., 71 percent in the U.K., 65 percent in Germany, and 59 percent in Australia. These estimates underscore that cross-country comparisons remain highly dependent on survey design, particularly on firm-sample coverage and on who is being surveyed.

China provides an especially relevant setting for our study. The 2024 National Economic Census collected information on enterprise AI use for the first time and reports an adoption rate of approximately 20 percent. The 2025 Enterprise Survey for Innovation and Entrepreneurship in China (ESIEC), which primarily covers micro and small firms, estimates that 26 percent of surveyed firm owners used AI in business operations (Song, Yang, and Zhang 2025). These differences are informative rather than merely contradictory: they suggest that the estimated prevalence of AI adoption in China is sensitive to the inclusion of small firms, the scope of AI use covered by the questionnaire, and the extent to which respondents can observe technology deployment within their organizations.

Reviewing the existing literature suggests that differences in reported AI adoption rates across

[7] At the individual worker level, Bick, Blandin, and Deming (2025) finds that, by late 2024, nearly 40 percent had used GenAI, and 23 percent of employed respondents had used it for work. These worker-level estimates capture individual task use rather than formal firm adoption, as the latter requires organizational shifts or investment.

surveys mainly stem from three sources. First, as AI adoption rates tend to be positively associated with firm size, [8] both the sampling frame and the weighting scheme applied to survey responses substantially affect the resulting estimate. In the U.S., the BTOS mirrors the actual firm population, where 95 percent of firms have fewer than 50 employees, and produces a firm-weighted estimate of 18 percent as of year-end 2025 (Allen 2026). The SBU oversamples large firms and reports an employment-weighted estimate, reaching a much higher rate of 78 percent (Yotzov et al. 2026). Second, how a survey question is framed substantially affects reported adoption rates. For instance, the Census Bureau revised its BTOS question in November 2025, which broadened the scope from AI use in "producing goods or services" to use in "any of its business functions." This change alone produced a discrete upward jump in reported adoption rates (Allen 2026; U.S. Census Bureau 2025). Similarly, when the OECD restricts its question to AI use in core business functions, estimated adoption rates fall far below the headline figures from broader national surveys (OECD 2025). The Eurostat survey, by contrast, broadened the range of specific AI technologies covered, yielding higher rates. Third, the composition of survey respondents and their incentives also matter. On the one hand, firm representatives may overreport adoption due to social desirability pressure as AI becomes a marker of organizational competitiveness.[9] On the other hand, operational staff may underreport due to limited visibility into firm-wide deployment (Allen 2026). Surveying senior executives partly mitigates this issue but does not fully resolve it (Yotzov et al. 2026).[10]

2.2 Patent-based approach

A second strand of research measures firms' AI-related innovation using patents, trademarks, or related intellectual-property records (e.g., Cheng, Lee, and Tambe 2023). Such measures capture costly and verifiable activities tied to the development or commercialization of AI, making them well-suited for studying frontier innovation or the geography of AI invention. They are less suitable for measuring adoption and diffusion: A firm can adopt a commercial large language model

[8] Acemoglu et al. (2022), McElheran et al. (2024), Eurostat (2025), and Song, Yang, and Zhang (2025).

[9] The general survey methodology literature establishes that respondents systematically misreport on topics where social norms create pressure to answer in a particular direction (Tourangeau and Yan, 2007; Groves et al., 2009). Allen (2026) and Yotzov et al. (2026) flag this as a plausible concern in the AI adoption context, though direct empirical evidence remains limited.

[10] See, also, the executive surveys used by Hoffreumon et al. (2023) and Fontanelli et al. (2024).

or reorganize workflows around GenAI tools without filing a single patent; conversely, a firm holding AI patents may develop technologies for sale to others without deploying them internally. Patent measures thus identify AI innovators and technology suppliers rather than the broader population of firms integrating AI into their operations. The distinction is visible in Calvino et al. (2022), who combine intellectual-property records, corporate websites, online job postings, and firm financial data to identify AI-active firms in the United Kingdom.[11]

2.3 Self-disclosure approach

Another approach measures firms' AI engagement by constructing text-based indicators from AI-related keywords or passages using public disclosures, such as annual reports, regulatory filings, earnings communications, and corporate websites. For example, Yao et al. (2024) measure AI-related keyword frequencies in the annual reports of Chinese listed firms. Similarly, Calvino et al. (2022) use AI-related content on corporate websites to identify AI-active firms in the United Kingdom. As these sources are widely available and readily linked to firm outcomes, they are useful for studying firms' strategic attention to AI and its relationship with financial performance or market valuation. Their primary limitation is that firms may discuss AI to signal technological orientation to investors or describe future strategies without any meaningful internal deployment or adoption.[12] Disclosure-based measures are therefore better interpreted as indicators of AI-related strategic orientation than as direct measures of realized adoption.

2.4 Job posting approach

A growing literature uses online job postings to measure firms' demand for AI-related capabilities (Babina et al. 2024; Hampole et al. 2025; Hosseini and Lichtinger 2025). This approach identifies vacancies requiring AI-related skills, occupations, or task knowledge, and then aggregates these postings to the firm, industry, or regional level. Relative to patents and public disclosures, job postings offer an appealing behavioral interpretation: recruiting workers with AI-related capabil-

[11] Their classification distinguishes firms conducting "AI innovation," firms with an "AI core business," and firms demanding "AI talent." This multidimensional framework is useful precisely because the three groups need not coincide. Firms developing AI technologies, firms publicly positioning themselves around AI, and firms recruiting AI-related human capital represent different forms of engagement with the technology.

[12] Moreover, keyword-based indices are difficult to interpret as adoption rates: more frequent mentions of AI do not necessarily imply more intensive use of the technology.

ities is a costly, observable organizational decision. Firms generally post vacancies to fill specific operational needs, and job advertisements reveal both the functions in which firms seek AI capabilities and the complementary skills that they regard as valuable.

Existing studies demonstrate the usefulness of this approach for measuring AI-related organizational investment. Calvino et al. (2022) construct an AI-hiring intensity measure for British firms, defined as the ratio of AI-related vacancies to total vacancies. They interpret this measure as a proxy for firms' specialization in AI and their longer-run commitment to AI capability building. Using U.S. data, Alekseeva et al. (2025) identify AI-related vacancies based on the co-occurrence of listed skills with unambiguous AI skills and construct a firm-year AI adoption proxy. More recently, Shi et al. (2025) focus directly on the measurement of AI jobs in online vacancies. They compare existing skill-taxonomy methods with fine-tuned large language model classifiers and construct a quarterly firm-level dataset for Russell 1000 firms. Their measure of AI Jobs Intensity, the ratio of AI-related job postings to all job postings, increases from 0.2 percent in 2018 Q1 to 0.8 percent in 2025 Q3. Importantly, this quantity measures labor demand for technical AI capabilities rather than the fraction of firms that have fully adopted AI in operational use.

Job-posting data offer several advantages for studying firm-level AI diffusion. First, they provide high-frequency, firm-level measures that can capture changes in AI-related organizational investment over time. Second, they reveal where in the organization AI capabilities are demanded, allowing researchers to distinguish technical positions from non-technical ones. Third, job postings are less likely to serve as mere rhetorical signals, as posting a vacancy and screening candidates require genuine organizational resources. This measure should nevertheless be interpreted with care. Firms may adopt off-the-shelf GenAI tools, rely on external providers, or encourage existing employees to use GenAI without posting new vacancies. Conversely, a GenAI-related vacancy indicates an intended investment that may not immediately translate into realized use.

3 Applying the Job Posting Approach to Chinese Firms

Given the advantages of the job-posting approach summarized above, we apply it to Chinese firms to measure the rate and intensity of GenAI adoption, and to identify the business functions in which these capabilities are demanded.

This section outlines the data sources and the steps we followed to build the firm-level AI-adoption dataset. We begin with an overview of job posting data (Section 3.1) and how they are merged with the firm-level data (Section 3.2). We then explain the deduplication process (Section 3.3), especially when postings are gathered from multiple job boards, and the keyword approach used to identify postings that require AI-related skills (Section 3.4). These postings are then manually reviewed to correct misclassifications and potential false positives (Type I errors). Then, an LLM-based classification method is employed to categorize AI-related posts by business function, leveraging details from job descriptions (Section 3.5). Finally, we present some emerging trends from the identified postings (Section 3.6).

3.1 Job Posting Data

Our job posting data come from three major online job boards in China: *51job.com*, *Zhaopin.com*, and *Liepin.com*, covering the period 2015–2025.[13] We have data access to all vacancies posted on *51job.com*, comprising 67 million unique postings over 2015–2025, and we restrict the sample to the 63 million postings matched to the business registry database (Section 3.2), covering approximately 2.5 million firms. From the firm registry, we can further identify about 6 million job vacancies posted by firms listed on China's stock exchanges.[14] The remaining 57 million jobs are posted by "non-listed" firms.

For the listed firms, we complement *51job.com* postings with those from the other two job boards, *Zhaopin.com* and *Liepin.com*, with stock IDs already labeled for matching. Pooling the three boards gives 16.9 million raw postings for listed firms. Because a firm can advertise the same vacancy on more than one platform, we deduplicate these cross-platform repeats before analysis (Section 3.3). After deduplication, we have 14.7 million unique postings for listed firms. The job posting data, therefore, support two samples that differ in platform coverage: a listed-firm sample

[13] *51job.com* was founded in Shanghai in 1998 and listed on NASDAQ in 2004. It had accumulated over 155 million registered users by 2021, with monthly active users reaching approximately 20 million as of December 2022. *Zhaopin.com* was founded in Beijing in 1994 and consistently ranks among the top two online recruitment platforms in China by monthly active users. *Liepin.com* was founded in 2011 and listed on the Hong Kong Stock Exchange in 2018. It differentiates itself by focusing on mid-to-senior white-collar and executive talent and has over 84 million registered users as of 2022. According to the *2024 China Online Recruitment Industry Research Report* conducted by iResearch (see https://report.iresearch.cn/report/202204/3981.shtml), online recruitment accounts for approximately 70% of the total recruitment activities, and these three job boards accounted for more than two-thirds of China's online recruitment market during most of our sample period.

[14] The term "listed firms" refers to firms publicly traded on mainland China's stock exchanges, including both the Shanghai Stock Exchange and the Shenzhen Stock Exchange, together known as the A-share market.

built from all three boards, and a non-listed-firm sample built from *51job.com* alone. In total, our sample contains 72 million unique postings for listed and non-listed firms together.

Each job posting contains structured information on the employer name, job title, occupational category, posting date, work location, educational requirements, number of vacancies, and salary, as well as unstructured text describing job responsibilities and skill requirements. The unstructured job description is typically several hundred Chinese characters long. We apply keyword matching and Large Language Model (LLM) methods to job titles and descriptions to determine whether AI-related skills are required (Section 3.4) and to identify the business functions that the worker is expected to support (Section 3.5).

Job posting data have certain limitations. Online job postings tend to provide better coverage of high-technology industries and high-skill occupations. Firms' decisions to post vacancies online may be selective and not perfectly aligned with their underlying labor demand (Kureková, Beblavý, and Thum-Thysen 2015). Nevertheless, compared with traditional surveys, job postings offer large-scale, timely, and detailed information on job requirements, skill demand, and geographic distribution, making them especially useful for tracking firms' hiring-based responses to fast-changing technologies such as GenAI.

3.2 Firm-level Data

We match job postings to two sets of firm-level data: 1) annual reports of listed firms, and 2) the business registry database for non-listed firms.[15]

Listed firms. We compile firm-year panel data from the annual financial reports of Chinese listed firms.[16] The dataset is sourced from Wind Information, containing up to 5,182 firms by the end of 2025.[17] These firms are classified into four broad sectors, namely manufacturing, mining and

[15]The two Chinese sources are 上市公司年报 (annual reports of listed firms) and 中国企业工商注册信息 (the business registry database).

[16]In China, listed firms are required to disclose their financial information and workforce structure annually. According to the *Content and Format Standards for Information Disclosure of Companies Issuing Securities Publicly No. 2—Content and Format of Annual Reports*, issued by the China Securities Regulatory Commission (CSRC), listed firms are required to disclose employment-related information. As stipulated in Article 34: "Firms shall disclose information about employees... including the total number of employees at the end of the reporting period, along with their functional composition (such as production, sales, technical, financial, and administrative personnel) and educational background..." These disclosures list employee numbers by function—such as sales, technical, production, administrative, and financial—offering insight into the firm's organizational structure.

[17]Link of Wind Information: https://www.wind.com.cn/.

construction, market services, and non-market services, and 18 more granular industries. Manufacturing firms account for 68 percent of the listed-firm sample. Most of the remaining firms operate in services, with market services accounting for 21 percent of the sample, including 8 percent in ICT and 2 percent in scientific research and technical services. As for ownership, 26 percent of listed firms are state-owned enterprises (SOEs),[18] which tend to be larger in employment size.

We link listed firms to job postings using stock IDs, which helps avoid matching errors arising from variations in firm names. As shown in Table A4, the matching rate increases from 87 percent in 2015 to 95 percent in 2025, with an average of above 90 percent over the sample period. This confirms that our job posting data provide comprehensive coverage of listed firms.

Non-listed firms. To identify and characterize non-listed firms, we match job postings to the State Administration for Industry and Commerce (SAIC) business registration database.[19] This database covers the universe of legally registered business entities in China and provides administrative records on registered name, industry classification, address, establishment and closure dates, registered capital, legal representative, and ownership structure. These records allow us to characterize firms by sector, location, age, operating status, and ownership type. They are particularly valuable for studying non-listed firms, which are typically absent from standardized financial databases and often lack consistent identifiers in job posting data.

We match job postings to the business registration database using standardized employer names, constructed by removing punctuation, whitespace, and common legal suffixes from firm names in both datasets.[20] 93 percent of the 67 million unique job postings from *51job.com* are successfully matched to registered business entities in the SAIC database, and in total, we match 63 million postings to approximately 2.5 million firms. Among them, 57 million job postings are from non-listed firms.

[18] We identify SOEs based on ultimate control: if the ultimate controlling shareholder is the government, the firm is classified as a state-owned enterprise.

[19] The SAIC business registration database has been widely used in the literature (Cong, Yang, and Zhang 2024; Dai and Zhang 2026). See Appendix A.1 for more details.

[20] When multiple registration records share the same standardized name, we use location and sector information as tiebreakers.

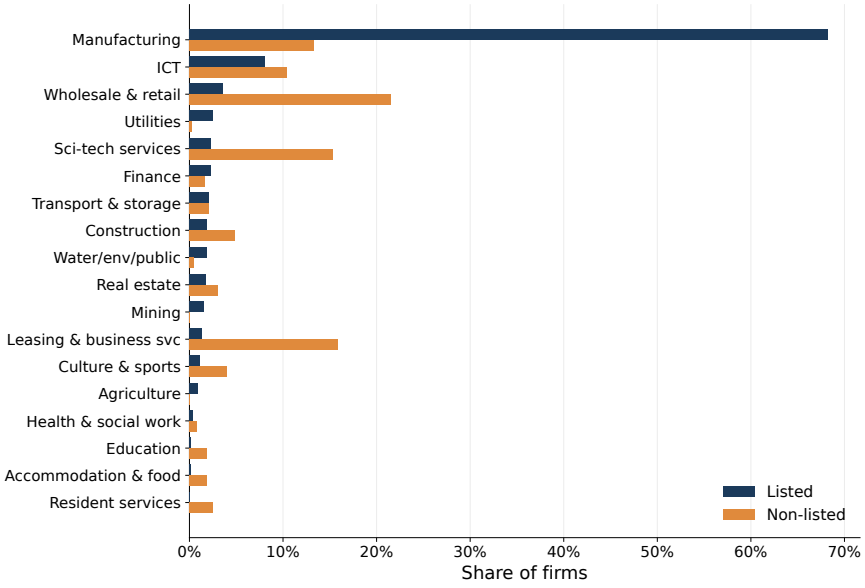


Figure 1: Sectoral Distribution of Listed and Non-listed Firms

Note: The figure shows how listed and non-listed firms are distributed across the industry sections of China's GB/T 4754 classification.

Listed vs non-listed firms Listed firms tend to be larger, more productive, and more technologically advanced than non-listed firms. In our matched sample, listed and non-listed firms also differ sharply in their sectoral composition (Figure 1). Listed firms are highly concentrated in manufacturing, which accounts for 68% of all listed firms, with very limited representation in agriculture, health and social work, and residential services. By contrast, non-listed firms are more evenly distributed across sectors and have a stronger presence in services, led by wholesale and retail trade at 22%, leasing and business services at 16%, and scientific and technical services at 15%. Manufacturing accounts for only 13% of non-listed firms. Sectoral composition and uneven GenAI exposure across industries may shape differences between listed and non-listed firms, which matters for interpreting the adoption patterns in Section 4.

3.3 Deduplication of Job Postings

For listed firms, our job postings are sourced from three online job boards, and a firm may post the same position on multiple boards. To prevent duplicates from influencing our results, we developed a deduplication algorithm for postings across different platforms.[21]

[21] Our approach is closely aligned with the deduplication method adopted by Revelio Labs, a workforce intelligence company (https://www.reveliolabs.com/). The differences are that we use a wider window to detect duplicates and

Deduplication Algorithm. The deduplication algorithm processes raw postings in three stages. We first extract and standardize key fields (location, job title, job responsibilities, and posting date); we then remove exact repeats within a short time window (*strict deduplication*); and we finally use a text-similarity measure across platforms to catch near-duplicates that differ only in formatting or wording (*loose deduplication*).

Extracting the job responsibility text is the main challenge, because raw descriptions of postings are highly unstructured.[22] We use a segment-based rule that captures the text following job-responsibility headers (e.g. *tasks*, *duties and responsibilities*) up to the next header of qualifications, skills, or company information.[23] And we then standardize the result by removing platform-specific formatting. *Strict deduplication* then removes identical postings defined by the same firm, the same city and county, the same standardized job title and raw description, posted within overlapping $\pm$3-day windows. This stage runs across platforms, capturing the same vacancy posted on multiple boards.

Loose deduplication catches near-duplicates that strict matching misses. For example, postings that share the same job responsibilities may have job titles with small differences on two platforms, such as "software engineer" vs. "software development engineer". For every posting pair from the same firm and city posted within three days,[24] We compute the Jaccard similarity of the standardized titles and of the responsibility sections,

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|},$$

where A and B are the sets of unique 2-grams from each title (or standardized job responsibilities). Posting pairs are flagged as duplicates if the Jaccard similarity of their standardized titles exceeds 0.8, or the similarity of their responsibility sections exceeds 0.6, with the latter threshold being lower due to the greater average length of the "job responsibility" section in a posting. Similarity

more granular location information to detect duplicates, so our deduplication method is stricter than Revelio Labs'. Moreover, we use Jaccard similarity to capture the similarity between two postings, as it is better suited to the Chinese-language context, where job postings are frequently re-posted with identical content in a different sequence, posting length varies substantially across firm types, etc. See more details in Appendix B.

[22] The responsibilities section may appear in no fixed order, recur multiple times, or be interspersed with unrelated material such as company background.

[23] When both a start and an end header are present, we keep the intervening text; with only one header we keep the text on the relevant side; with none we fall back to the full description. Extractions shorter than 15 characters are treated as failures, and multiple qualifying segments are concatenated. See Appendix B.

[24] Pairs differing only in county-level location are kept as distinct vacancies.

is computed at the character level rather than at the word level, as Chinese text lacks whitespace delimiters and character bigrams capture sub-word overlap more reliably in this setting. Flagged pairs are then clustered, and only the earliest-dated posting is retained in one cluster.[25]

Deduplication results. The raw data contain 16.9 million postings in total, of which 2.2 million (13 percent) are identified as duplicates and removed. The duplicate share is relatively stable across years, ranging from 9% in 2015 to 15 percent in 2022, suggesting that cross-platform re-posting is a persistent rather than episodic feature of the data. Annual posting numbers follow an inverted-U pattern, peaking in 2022 and then declining through 2024-2025, consistent with the broader employment contraction in China's labor market in recent years. After deduplication, the cleaned sample retains 14.7 million postings, which forms the basis for all subsequent analysis.[26]

3.4 The Keyword Approach

We construct three keyword categories to measure (Gen)AI adoption: tools associated with GenAI applications ("*GenAI tool*"), technical terminology specific to GenAI ("*GenAI tech*"), and AI terms that predate the emergence of GenAI ("*Earlier AI*"). This taxonomy reflects a meaningful skill hierarchy. *GenAI tool* keywords capture entry-level, cross-functional GenAI skills relevant to a broad range of firms and industries, while *GenAI tech* keywords capture more specialized competencies typically concentrated in technical functions such as research and development. The *Earlier AI* category identifies earlier waves of AI adoption and serves as a benchmark.

A firm is classified as a GenAI adopter upon posting its first vacancy explicitly requiring either *GenAI tool* or *GenAI tech* skills.[27] To capture the intensive margin of adoption, we also construct a continuous measure defined as the share of a firm's postings that list GenAI skill requirements.

Compiling the lists. The three keyword lists are compiled systematically and tailored to the Chinese context (see Appendix Table A1 for the complete lists). The first category, *GenAI tool*, encompasses specific GenAI products that employers explicitly require workers to use, ranging

[25] As we receive the full set of postings from *51job.com*, we apply *strict deduplication* to all of its postings, which yields 67 million unique postings spanning both listed and non-listed firms (down from 77 million raw postings). The 57 million postings from non-listed firms, after merging with business registry data, all come from this single platform, so they do not require cross-board (*loose*) deduplication.

[26] Table A5 presents the number of job postings before and after deduplication across the sample period.

[27] See, for instance, Babina et al. (2024), Liu and Webber (2026), Alekseeva et al. (2020), and Goldfarb, He, and Teodoridis (2025).

from internationally prominent platforms, such as ChatGPT, Gemini, and Claude, to domestically developed models, such as DeepSeek, Qwen, Ernie Bot, and Doubao. The second category, *GenAI tech*, includes more technical terms such as transformers, fine-tuning, retrieval-augmented generation (RAG), and prompt engineering. It is worth noting that some of the GenAI technical terms, such as Transformers, existed before GenAI tools became widely available (i.e., around the end of 2022).[28] The third category, *Earlier AI*, covers machine learning and artificial intelligence technical terms that predate ChatGPT's launch in November 2022, such as deep learning, computer vision, and natural language processing (NLP).

The lists are developed from considering industry terminology, prior literature, and manual inspection of job postings to ensure broad coverage and relevance in the context of China. For instance, the construction of the GenAI application tool list begins with 60 top global GenAI applications in Liu, Huang, and Wang (2025) and Liu and Wang (2026), augmented with tools from analogous rankings for the Chinese market. All three lists are tailored to the Chinese context, both by emphasizing tools with substantial domestic market penetration and by including terms in both English and Chinese.[29] The three lists are applied to the 72 million job postings in our sample to identify postings that include any of these keywords in their job descriptions, particularly in the "job responsibilities" section. In total, we identified approximately 0.5 million job postings with related keywords.

Validation. After identifying postings containing the relevant keywords, we randomly sample a few thousand identified postings for manual review. This inspection reveals that certain AI-related technical terms are susceptible to conflation with unrelated concepts, potentially generating false positives (Type I errors).[30] Based on these insights from the sampled postings, we refine the keyword list and identification algorithm to reduce such Type I errors, which account for a negligible fraction of the identified postings.

[28] This is less true for GenAI tools, but there are a couple of cases, such as Midjourney, where GenAI tools were available before the launch of ChatGPT.

[29] Many technical terms appear in both languages within Chinese job postings, as practitioners frequently use English acronyms or brand names alongside their Chinese translations; matching on either form is therefore necessary to avoid systematic undercounting. Where no widely adopted Chinese equivalent exists, only the English form is matched; where a Chinese expression has no standard English counterpart, only the Chinese form is used.

[30] For example, the acronym *LLM* may denote either a large language model or a Master of Laws degree, and *GAN* may refer to either a Generative Adversarial Network or a Romanized transliteration of a Chinese proper noun.

3.5 Classifying Postings by Business Functions

To better understand AI adoption across business functions within firms, we develop a business function classification algorithm, taking into account the Chinese business context. We define five functional categories, namely technical, production, sales, administrative, and financial, to match the information in the financial report data for listed firms.

The classification approach proceeds through a cascaded three-tier architecture, combining rule-based matching, semantic embeddings, and LLM reasoning. Tier I applies deterministic keyword matching to job titles. For each of the five functional categories, we compile a list of high-frequency, unambiguous title keywords drawn from Chinese labor markets. For instance, titles containing terms such as "Sales" or "Marketing" are assigned to the Sales function. In Tier II, titles that do not match any Tier I keyword are passed to semantic vector similarity, which captures roles that share functional meaning with the keyword list but lack an exact lexical match. We employ the *text2vec-base-chinese* sentence embedding model, a transformer-based model pre-trained on large-scale Chinese corpora that encodes semantic relationships among words and phrases in Chinese text. In Tier III, the most ambiguous cases are classified using the Qwen-Plus model, prompted with a structured expert system. The expert system encodes category definitions, boundary rules, and disambiguation logic for high-frequency ambiguous terms.[31]

3.6 Emerging Trends from the Identified Postings

As illustrated in Figure 2, several keyword trends have emerged from the data. First, the share of postings requiring GenAI tools has risen sharply since 2022, reaching 0.25 percent among listed firms and 0.09 percent among non-listed firms by the end of 2025. About 65 percent of this increase occurred in 2025, driven by rising demand for DeepSeek. Although ChatGPT remains the most popular GenAI tool, firms increasingly prefer domestic ones such as DeepSeek, Qwen, and Doubao. The pattern is evident in both listed and non-listed firms, with non-listed firms showing a greater reliance on domestic tools.

[31] See Appendix Section C for more details.

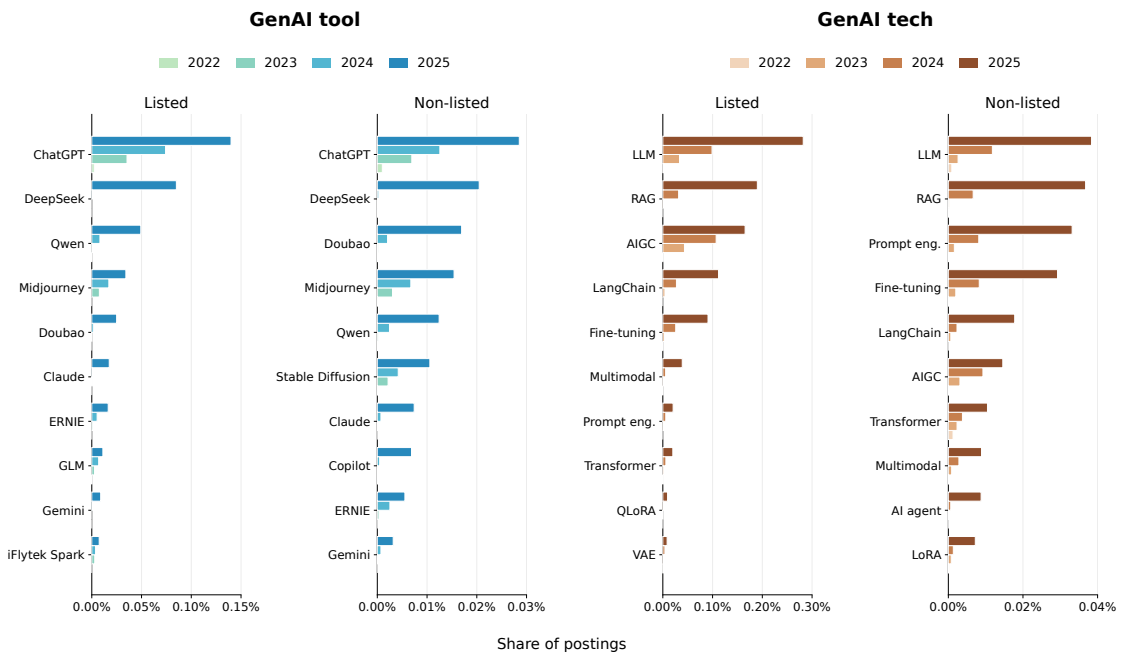


Figure 2: GenAI Keyword Trends, % of Total Postings

Note: Each panel shows the top ten keywords in a domain (GenAI tool, GenAI technique) for one type of firm (listed, non-listed). In 2025, the ten most frequent GenAI tools account for 96.7% of all GenAI-tool postings among listed firms and 94.9% among non-listed firms; the ten most frequent GenAI-technique terms account for 99.1% and 93.5%. Keywords that refer to the same tool or technique, including English and Chinese variants, are merged into a single item (e.g., ChatGPT/GPT; GLM/Zhipu; LLM/large language model).

Among all the GenAI technical terms, AIGC (AI-Generated Content) remained the most popular term until 2025, driven by rising demand for LLM and RAG (Retrieval-Augmented Generation)-related skills. By 2025, postings mentioning LLM and RAG had overtaken AIGC. This shift may indicate that the purpose of AI adoption is moving from simpler applications (like text or image generation) to more sophisticated, productive uses, as RAG is typically used to connect AI models to firms' proprietary, domain-specific data. This allows models to answer highly specific questions using internal documents, reduce "hallucinations," and provide verified citations without the high costs of retraining the underlying model.

Figure 3 shows how the intensity of posting requiring GenAI-related skills has evolved. Overall, the share of postings requiring GenAI-related skills (either tools or technical terms) remains small but has been rising rapidly. By the end of 2025, this share has reached 0.7 percent of total postings made by listed firms and 0.2 percent among non-listed firms, up from near zero by the end of 2022. In both listed and non-listed firms, the increase in demand for workers with GenAI skills has intensified over time, with the 2025 surge driven more by demand for workers with

GenAI technical skills. The increase in demand is more pronounced in listed firms, resulting in a widening gap between listed and non-listed firms (i.e., from a near-zero gap in 2022 to a gap of about 0.5 percentage points in 2025, see the right panel of Figure 3).

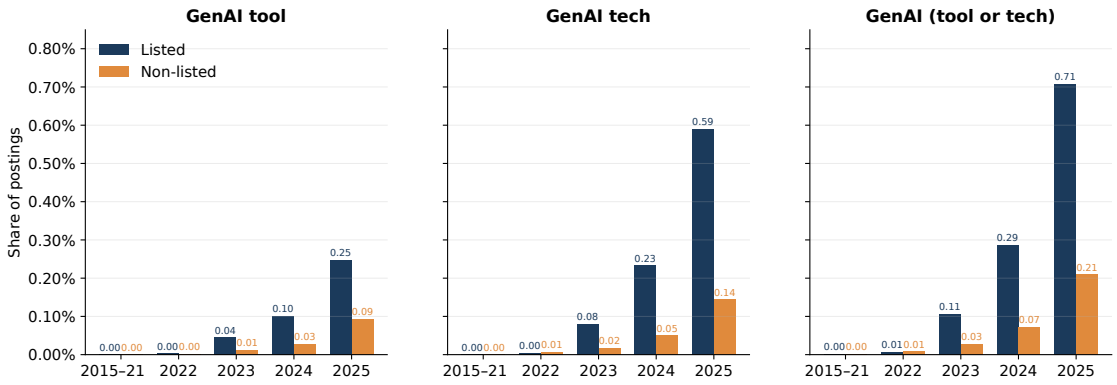


Figure 3: GenAI Job Postings, % of Total Postings

Note: The figure shows the share of job postings that require a GenAI skill, by type (GenAI tool, GenAI technique, and the union of the two), for listed and non-listed firms. Each panel pools 2015–2021 as a near-zero baseline and then reports 2022 through 2025.

Figure 4 compares the diffusion of "earlier" AI technology against generative AI using posting intensity. On average, the share of postings requiring earlier AI technical skills was about 0.7 (0.3) percent among listed (non-listed) firms over the period 2015-2021. The rise of GenAI around late 2022 appears to have increased demand for workers with earlier AI skills, with the share rising to 1.7 percent for listed firms and 0.7 percent for non-listed firms.[32] This seems to suggest that these earlier AI skills may complement GenAI skills rather than replace them. Notably, demand growth for earlier AI skills is more pronounced among listed firms, widening the adoption gap between listed and non-listed firms. When considering both earlier AI and GenAI-related skills together, over 2 percent of postings in listed firms and 0.8 percent in non-listed firms required any AI skills. Although listed firms had roughly twice the posting intensity for any AI skills during 2015-21, this increased to 2.5 times by late 2025.

[32] A similar pattern was also found in the case of U.S. (Kalyani et al. 2025).

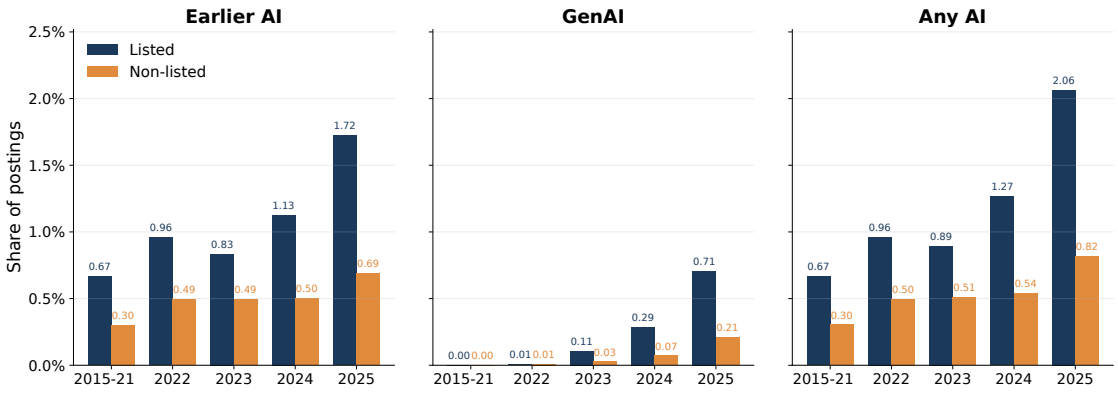


Figure 4: Posting-Level Adoption, by Measure

Note: Share of job postings that require an earlier-AI, GenAI, or any-AI skill, by period, for listed and non-listed firms.

4 GenAI Adoption among Chinese Firms

This section first shows stylized facts on Chinese firms' GenAI adoption patterns, followed by some early evidence on new GenAI-related occupations. It then presents a section that empirically examines which firm features are linked to GenAI adoption.

4.1 Stylized Facts from the Job Posting Approach

Fact 1: Adoption rate. We begin by documenting the overall adoption rate over time. Figure 5 presents the cumulative GenAI adoption rate, by month, for listed and non-listed firms. A firm is counted as a GenAI adopter in the month it first posts a vacancy that requires a GenAI skill (including tool or technical skills). By the end of 2022, right after the release of ChatGPT, adoption remains near zero for both listed and non-listed firms. The adoption rate then picks up among listed firms, reaching 11 percent by the end of 2024 and 21 percent by the end of 2025. Non-listed firms adopt at a much lower rate. Only 0.2 percent of non-listed firms have recruited workers with some GenAI skills by the end of 2025.

As shown above, the adoption gap between listed and non-listed firms has been widening over time. Part of this gap reflects the fact that non-listed firms are smaller and less financially able to make employment adjustments or recruitment decisions to adopt the new technology. It is also possible that non-listed firms are less technically sophisticated than listed firms, so that GenAI

technologies do not deliver sufficient productivity gains to justify their costly adoption (such as hiring a new worker).

The adoption rates for both listed and non-listed firms also reveal a clear "DeepSeek moment" in which they rise sharply around the public release of DeepSeek in January 2025 (the second dashed line in Figure 5). Among listed firms, GenAI adoption has expanded rapidly after the release of DeepSeek, rising from 11 percent at the end of 2024 to 21 percent at the end of 2025. This is the largest single-year increase that has been recorded so far, much larger than what is seen after the release of ChatGPT. The surge was even sharper for non-listed firms, where the adoption rate rose 2.7 times between 2024 and 2025 (reaching 0.17 percent by the end of 2025).

The influence of DeepSeek is most visible in postings requiring GenAI application tools. Among non-listed firms, the share of firms that have had postings requiring such tools more than tripled in 2025, rising from 0.03 to 0.09 percent. Meanwhile, the share of GenAI tool adopters among listed firms doubled in 2025, from 7 to 14 percent. This finding suggests that by offering a domestic, low-cost, and readily localizable alternative to frontier models, DeepSeek plausibly reduced both price and integration barriers to adoption. As these barriers are most binding for smaller, less resourceful firms, a model like DeepSeek might have disproportionately broadened access at the lower end of the firm-size distribution (like non-listed firms).

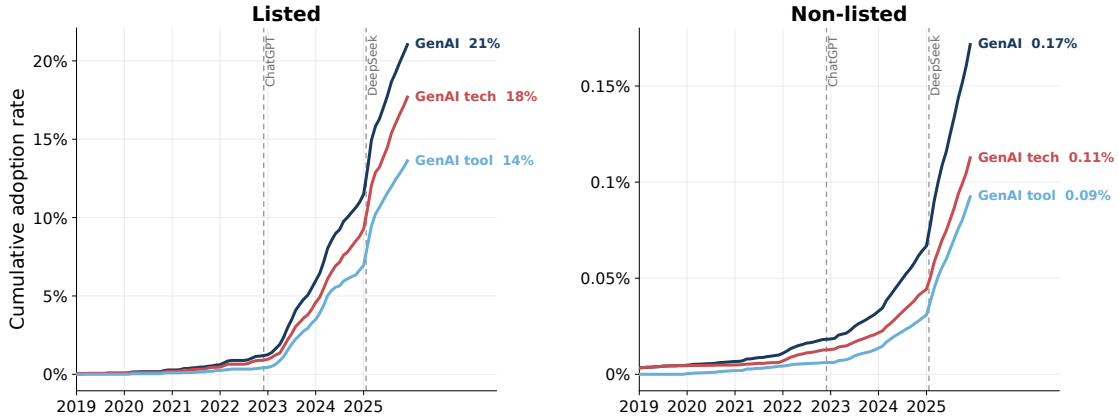


Figure 5: Cumulative GenAI Adoption Rate, 2019–2025

Note: Each line traces the cumulative share of firms that have adopted GenAI, by month, from January 2019 to December 2025. A firm counts as an adopter from the month of its first job posting that requires a GenAI-related skill. We plot three series: GenAI tool, GenAI tech, and the union of the two ("GenAI"). The denominator is the number of listed firms for the left panel, and the number of non-listed firms registered by that month for the right panel. The two dashed lines mark the public release of ChatGPT (November 2022) and DeepSeek (January 2025).

Comparing the diffusion speed of GenAI with earlier AI technologies shows that GenAI has

been adopted faster among listed firms (Figure 6). Earlier AI technologies, especially deep learning and advanced machine learning techniques, experienced a boom in the 2010s, driven by access to large amounts of data (known as "big data") and by cheaper, faster computers. In China, the adoption rate of earlier AI technologies among listed firms rose from 12 percent in 2015 to 23 percent in 2017 and to 56 percent in 2025. If we take 2010 as the starting point for earlier AI technology diffusion, we see that it took earlier AI technologies about 7 years (i.e., from 2010 to 2017) to reach a comparable level that took GenAI about three years to reach (i.e., from close to zero in 2022 to 21 percent in 2025). The rapid spread of GenAI could be attributed to the availability of Generative AI tools that are easier for workers with limited computer science backgrounds or basic ICT skills to adopt productively. The previous technologies are largely reserved for ICT engineers working on more technical functions.[33]

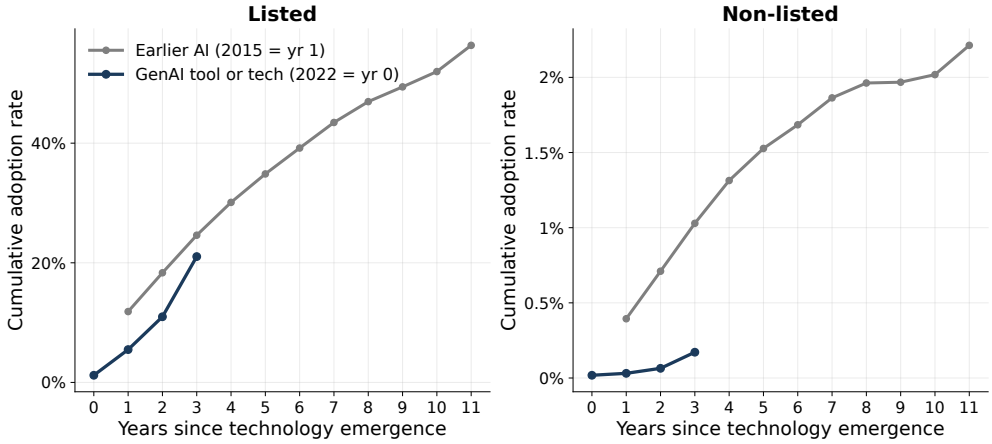


Figure 6: Diffusion of GenAI versus Earlier AI

Note: The figure compares how fast firms take up GenAI against how fast they took up earlier AI. Both series are placed on event time. For GenAI (tool or tech), year 0 is 2022, the year ChatGPT was released. For earlier AI, year 1 is 2015, the start of our posting data. The vertical axis is the cumulative share of firms that have posted at least one vacancy requiring the relevant skill. The left panel covers listed firms, and the right panel covers non-listed firms.

Meanwhile, the same surge in GenAI adoption was not observed among non-listed firms. The share of non-listed firms that adopted earlier AI technologies rose from below 0.5 percent in 2015 to above 2 percent in 2025. Adoption picked up notably between 2015 and 2022, then slowed after the rise of GenAI tools. The share of non-listed firms that adopted GenAI tools or techniques reached only 0.2 percent, one-fifth of the level observed in 2017. If this trend persists, the adoption gap between listed and non-listed firms may be wider for GenAI than for earlier AI technologies.

[33] Appendix D repeats the adoption patterns for earlier AI, GenAI, and any AI (their union) across the cuts used in Section 4.

The gap could be driven by the fact that listed firms are more likely to adopt GenAI by hiring new workers, while non-listed firms may adopt it narrowly by accessing GenAI applications. It is also possible that the high adoption cost of GenAI does not justify the benefits for non-listed firms, where scale effects may be limited.

Fact 2: Adoption by sector. GenAI adoption varies across sectors. Figure 7 shows the GenAI adoption rate by four broad sectors for listed and non-listed firms. Among listed firms, market services sectors have the highest adoption rate, reaching 35 percent by the end of 2025, ahead of non-market services at 23 percent, manufacturing at 17 percent, and construction, mining, and utilities at 15 percent. Non-listed firms show a different adoption pattern across the four broad sectors. The highest adoption rate was 0.3 percent among manufacturing firms, followed by 0.2 percent in market services. The adoption rates for non-listed firms in non-market services, construction, mining, and utilities were below 0.1 percent. The results are largely consistent with existing studies, which show that jobs in the market service sector are more exposed to Generative AI (Gmyrek, Viollaz, and Winkler 2026).

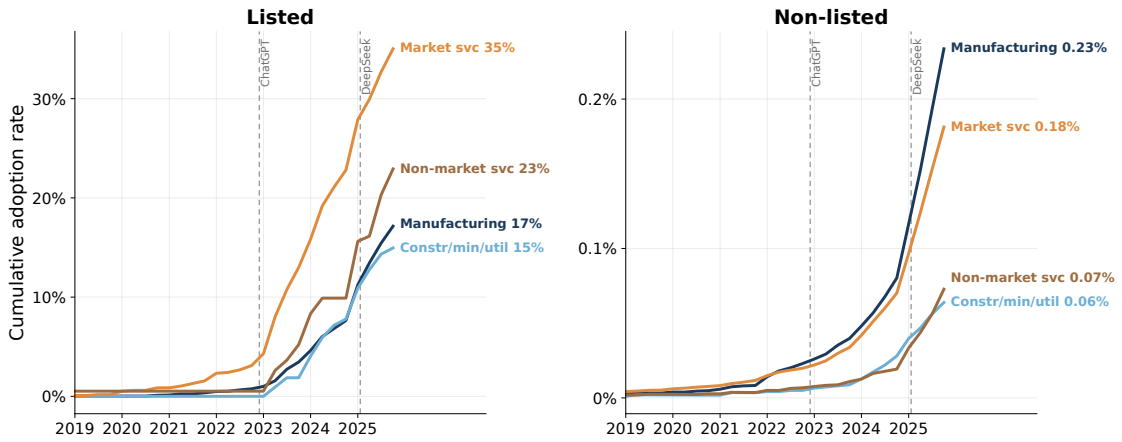


Figure 7: Cumulative GenAI Adoption Rate by Broad Sector, 2019–2025

Note: Each line traces the cumulative share of firms in a broad sector that have adopted GenAI (tool or tech), by quarter. Firms are grouped into four broad sectors: manufacturing; construction, mining, and utilities; market services; and non-market services. The denominator is the total number of firms existing in the sector-quarter cell. The left panel covers listed firms, the right panel non-listed firms.

The GenAI adoption rates also differ across industries within each broad sector (Figure 8). Among listed firms, the adoption rate is highest in information and communication technology (ICT) at 48 percent, education at 46 percent, and finance at 41 percent. It is lowest in accommodation and food, mining, and utilities. The non-listed firms showed a similar ranking, with the

highest adoption rates observed in the ICT industry, finance, and science and technology services, and the lowest among accommodation and food. Furthermore, the adoption pattern seems to differ between listed and non-listed firms for a few industries, such as education, manufacturing, utilities, and construction. These differences could stem from differences in firm size and the industry value chain segments in which they operate between the two groups of firms.

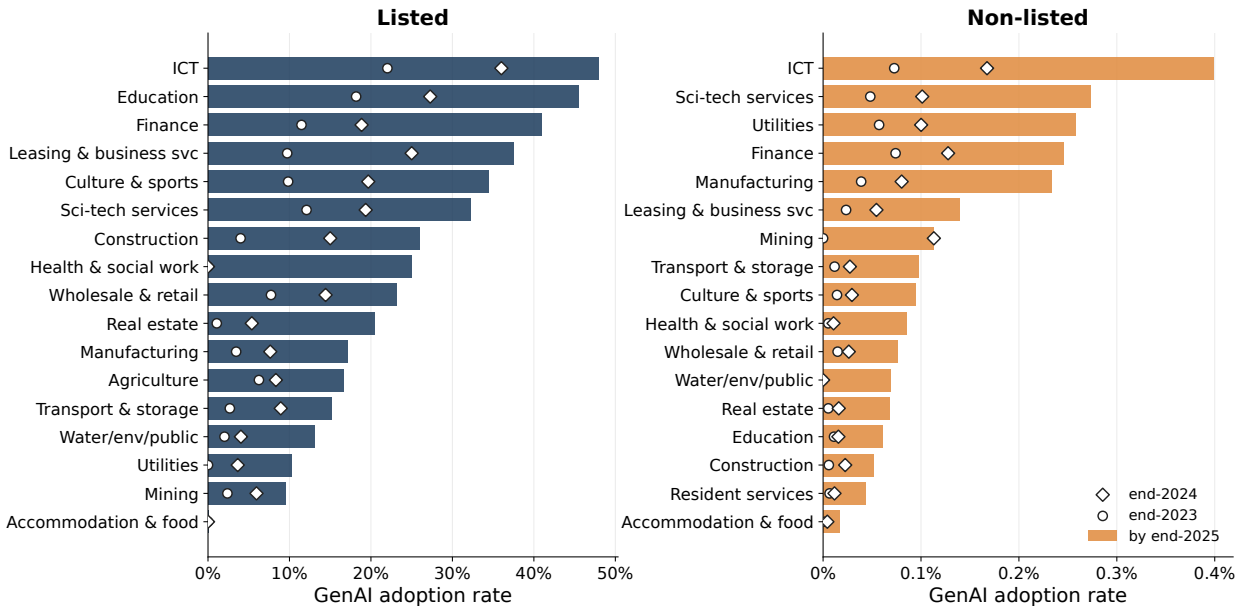


Figure 8: GenAI Adoption Rate by Industry, End of 2025

Note: Bars show the share of firms in each industry that have adopted GenAI by the end of 2025. Open circles mark the share by the end of 2024 and open diamonds the share by the end of 2023, so the gap between the bar and the markers shows recent growth. Industries follow the single-letter sections of China's GB/T 4754 classification; public administration is dropped, as is agriculture for non-listed firms, and we keep only industries with at least five firms. Each panel is sorted by its own 2025 rate.

Fact 3: Adoption by size. Here we plot firms' adoption rates by size. Since we do not have employment sizes for non-listed firms, we assume a positive relationship between employment size and posting number and measure firm size as the total number of vacancies a firm posts over the sample period, placing listed and non-listed firms on a common scale. As shown in Figure 9, the size distribution of GenAI adopters is notably more skewed to the right than that of non-adopters, suggesting that GenAI adopters tend to be larger in employment size than non-adopters. This pattern is observed in both listed and non-listed firms. The median listed adopter posts 3,378 vacancies over the sample period, against 681 for non-adopters. Among non-listed firms, the medians are 112 and 5. This finding aligns with existing studies that show larger business organizations are more likely to adopt AI and make AI-related hiring decisions (e.g., Kazakova

et al. 2020; McElheran et al., 2024; McKinsey, 2025).[34] It is possible that larger firms may have more resources to experiment with new technologies and can better absorb the fixed costs of AI adoption. Additionally, the scaling effects of AI might also reward bigger companies, motivating them to adopt.

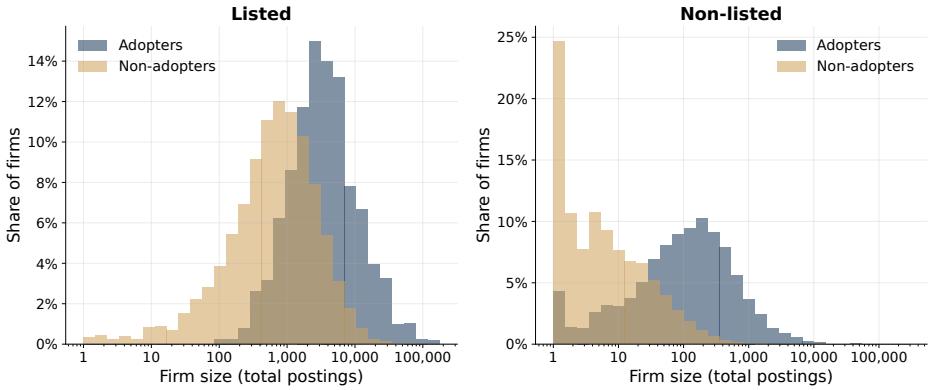


Figure 9: Firm Size of GenAI Adopters and Non-adopters

Note: The figure compares the size distribution of GenAI adopters and non-adopters. We measure firm size by the total number of job postings a firm makes over the whole sample, which lets us compare listed and non-listed firms on one scale; Appendix Figure A1 shows that this proxy lines up with reported employment for listed firms. The x-axis is on a log scale, and the y-axis is the share of firms in each size bin. The left panel covers listed firms, and the right panel non-listed firms. The median adopter posts 3,378 vacancies versus 681 for non-adopters among listed firms, and 112 versus 5 among non-listed firms.

The GenAI adoption rate also increases with firm size within each sector (Figure 10). The increase with size is steepest for market and non-market services, more modest for manufacturing, and almost flat for construction, mining, and utilities. Among listed firms, the adoption rate increases from about 1 percent in the smallest size bin to about 60 percent in the largest. Meanwhile, among non-listed firms, the adoption rate increases more steadily with size for manufacturing and market services but moderately for non-market services and construction, mining, and utility. The adoption rate rose from zero in the smallest size bin to about 1 percent in the largest size bin (i.e., close to the smallest size bin for listed firms). These patterns suggest that both sectoral features and firm size help shape GenAI adoption decisions.

[34] Former studies, such as Zolas et al. (2020), also showed that advanced technology adoption is generally skewed towards larger and older firms.

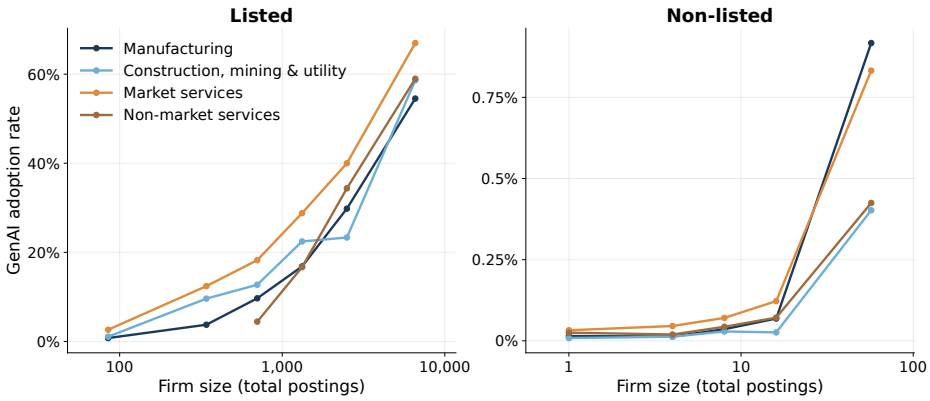


Figure 10: GenAI Adoption Rate by Firm Size and Sector

Note: The figure plots the GenAI adoption rate against firm size, separately for each broad sector. Firm size is the total number of postings over the sample, split into six equal-count bins; the x-axis is on a log scale. The left panel covers listed firms, and the right panel non-listed firms.

Fact 4: Adoption by ownership (for listed firms). For listed firms, we are able to sort them into state-owned enterprises (SOEs) vs non-SOEs and compare their GenAI adoption rates. Among listed firms, SOEs and non-SOEs adopted GenAI at a similar overall rate, 21 percent and 22 percent, respectively, by the end of 2025 (Figure 11). However, when we plot GenAI adoption rates by employment size, we notice that, at the same employment levels, there are more non-SOEs than SOEs that have adopted GenAI. For instance, among firms with about 12,000 employees, 47 percent of non-SOEs have adopted GenAI, against 38 percent of SOEs. The similar overall adoption level between the two firm groups is mainly due to the fact that listed SOEs are larger than non-SOEs. In addition, government policies may help SOEs keep the overall adoption rate. For instance, the State-owned Assets Supervision and Administration Commission (SASAC) has consistently encouraged State-Owned Enterprises (SOEs) to adopt AI.[35]

[35]For example, in February 2024, China's State-owned Assets Supervision and Administration Commission (SASAC) convened a special meeting to advance the integration of "AI+" initiatives into the operations of central state-owned enterprises.

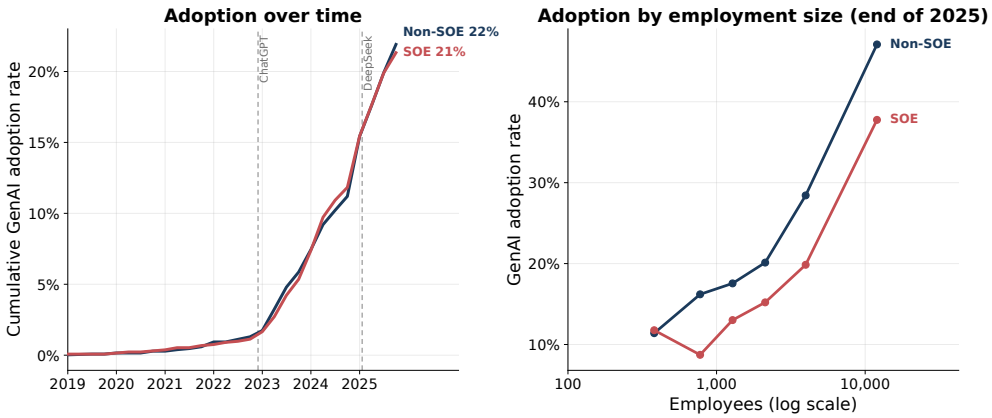


Figure 11: GenAI Adoption by Ownership Type, Listed Firms

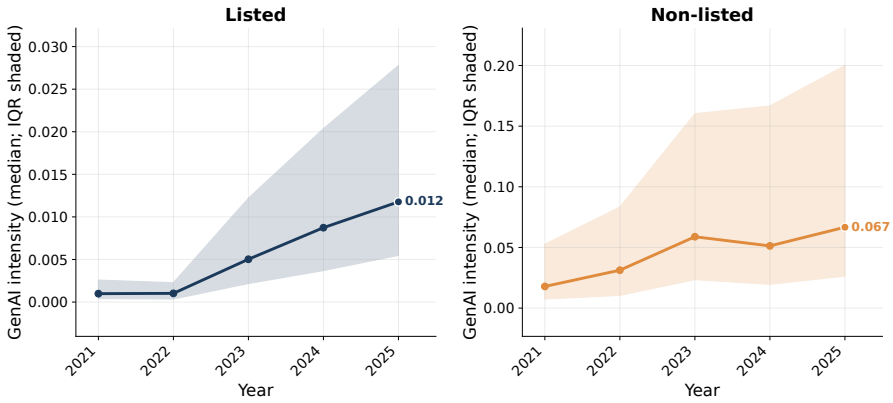
Note: The figure covers the listed firms only. The left panel plots the quarterly cumulative GenAI (tool or tech) adoption rate for state-owned enterprises (SOEs) and non-SOEs, with the denominator fixed at the number of firms in each group; the dashed lines mark the release of ChatGPT (November 2022) and DeepSeek (January 2025). The right panel plots the GenAI adoption rate by the end of 2025 against firm employment on a log scale, averaged within size bins, for SOEs and non-SOEs.

Fact 5: Adoption intensity. We then show how adoption intensity, measured as the share of a firm's postings that require (Gen)AI-related skills relative to total postings, evolves over time. Figure 12 presents the median and interquartile ranges of adoption intensity across all firms that adopted GenAI, with separate breakdowns by firms' listing status. Figure 12a shows the adoption intensity among firms that have at least one posting requiring GenAI tools or technical skills.[36] The lines show the median rates, and the shaded area represents the interquartiles. Across both listed and non-listed firms, the median intensity increases steadily over time, with the top quartile rising faster than the median and the bottom quartile. This indicates that, while overall adoption intensity is moderate, some firms are adopting GenAI more aggressively than the sample majority.

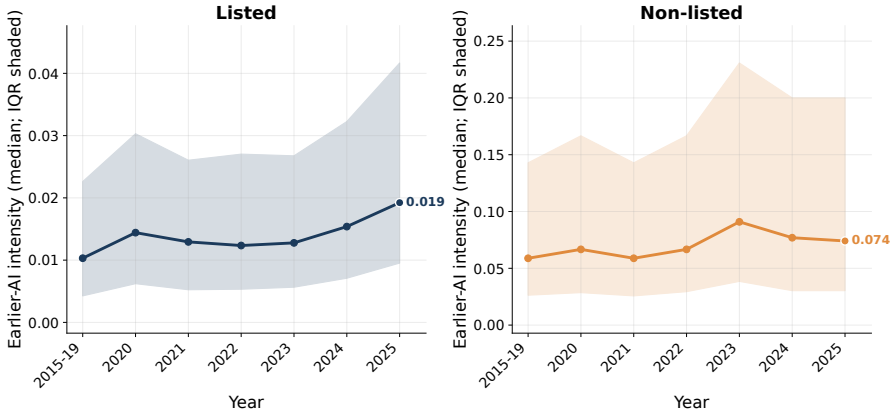
Figure 12b repeats the exercise for earlier AI adoption. As postings requiring earlier AI skills are common over the whole sample period, we also show the intensity levels before 2021, with the level averaged over the period 2015-19. We find that the intensity for earlier AI skills has remained largely stable over the past decade. While earlier AI intensity levels are notably higher than GenAI intensity in both firm groups at the beginning of 2021, they are comparable by the end of 2025. This is likely because earlier AI technologies have been around longer and have reached a mature stage. It is also possible that the rapid increase in adoption intensity for GenAI skills

[36] We begin in 2021 because postings requiring GenAI-related skills before then come from too few firms to summarize a representative sample. The full year-by-year distributions are in Appendix Figure **??**.

indicates that its adoption barriers might be lower than those for earlier AI technologies.



(a) GenAI (tool or tech) intensity



(b) Earlier AI intensity

Figure 12: Adoption Intensity: Median and Interquartile Range by Year

Note: Intensity is the share of a firm's postings in a year that require the relevant AI skill, computed among firms with at least one such posting (firm-years with none are dropped). In each panel the line is the median and the shaded band the interquartile range; listed firms are on the left and non-listed firms on the right, each on its own scale, and the label marks the 2025 median. Panel (a) covers GenAI (tool or tech) and begins in 2021, because GenAI postings before then come from too few firms to summarize a representative sample. Panel (b) covers earlier AI, which is common throughout the period, so it begins with a pooled 2015–2019 point followed by individual years. The full year-by-year distributions appear in Appendix Figures **??** and **??**.

Fact 6: Adoption sophistication. Here, we utilize the two different sets of GenAI keywords to understand the sophistication of GenAI adoption. We treat GenAI tech term adoption as more technically advanced than GenAI tool adoption, and split adopters into three levels: tool-only, tech-only, and both tool and tech. The results are shown in Figure 13. We find that listed firms are notably more likely to adopt both GenAI tools and technical skills, while the majority of non-listed firms have adopted either GenAI tools or technical skills. Moreover, the rate of GenAI tool adoption is much higher for non-listed firms, ranging from 33 percent for market services to 48

percent for non-market services, than for listed firms (i.e., 12 percent for market services and 27 percent for non-market services). The results also show that adoption tends to be more technically sophisticated in market services than in other sectors, across both listed and non-listed firms, as indicated by the highest level of GenAI tool and technical adoption.

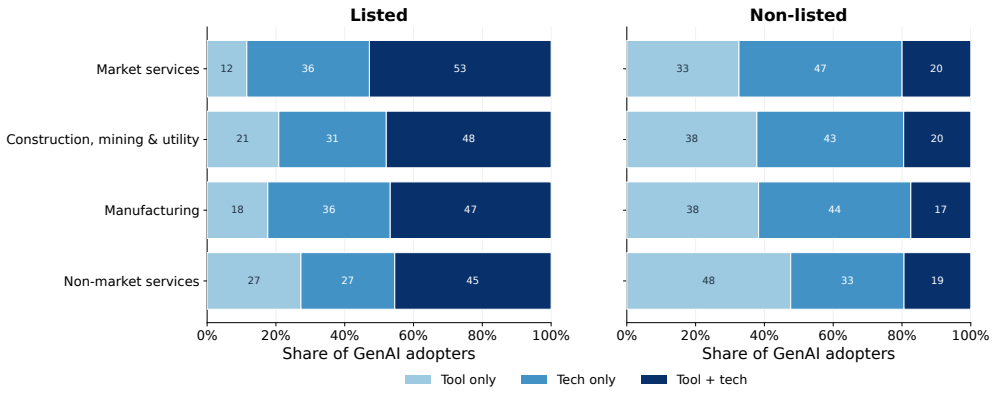


Figure 13: Sophistication of GenAI Adoption, by Sector

Note: Among GenAI-adopting firms, the bars decompose adoption into three rising levels of sophistication: tool only (lightest), tech only, and both tool and tech (darkest). Bars are 100 percent stacked within each broad sector and sorted by the share that uses both. The left panel covers listed firms, and the right panel non-listed firms.

Furthermore, we show which business functions listed firms adopt GenAI for, and how this varies across sectors. As mentioned above, when listed firms post vacancies online, they need to indicate which business functions they are recruiting for. As detailed in Section 3.5, we used job titles and descriptions about job responsibilities to map each GenAI-related posting to one of the five core business functions: production, sales, technical, administrative, and financial (see Appendix C for the detailed methodology). As shown by Figure 14, across all broad sectors, the adoption rate is highest for technical functions, followed by sales and administration, with the lowest adoption rate found for production and financial functions. The adoption rate for production, administrative, and financial functions each remains at or below 4 percent across all sectors. This indicates that only the most intensive adopters, concentrated in market services, extend GenAI beyond technical roles into customer-facing and back-office work.

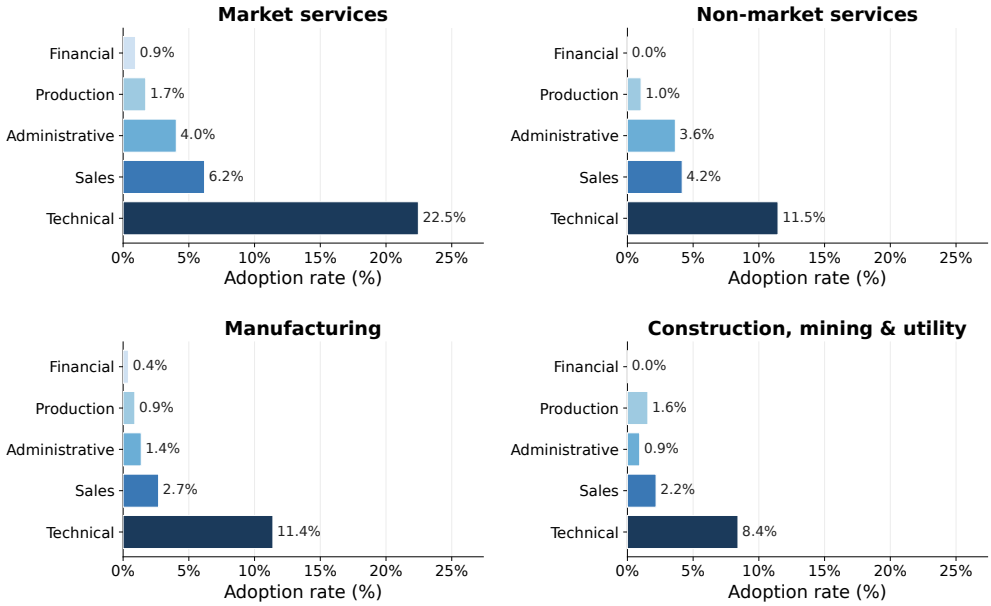


Figure 14: GenAI Adoption Rate by Business Function, within Each Sector (Listed Firms)

Note: Each panel covers one of the four broad sectors. A bar shows the share of listed firms in that sector that posted at least one GenAI vacancy in the corresponding business function. Functions are taken from the posting's job description field using the methodology documented in Section 3.5 and mapped to the five annual-report categories (more details see Appendix Section C).

Fact 7: Adoption and pre-existing skilled workers. Here, we examine whether firms that hired for earlier AI skills are more likely to adopt GenAI later. Figure 15 splits firms by whether they had posted an earlier-AI vacancy by 2022 and compares the two groups' GenAI adoption rates. Among listed firms, 37 percent of those that have previously adopted AI skills go on to adopt GenAI by the end of 2025, compared with 7 percent for the rest. The gap is just as sharp for non-listed firms, 2.9 percent among those that have adopted earlier AI skills versus 0.1 percent among those that have not. The results suggest that having a pool of AI-skilled workers lowers the cost of adopting GenAI (new AI technologies).

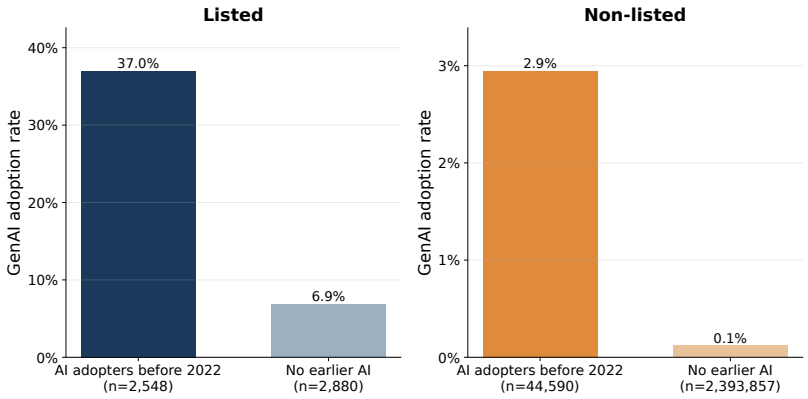
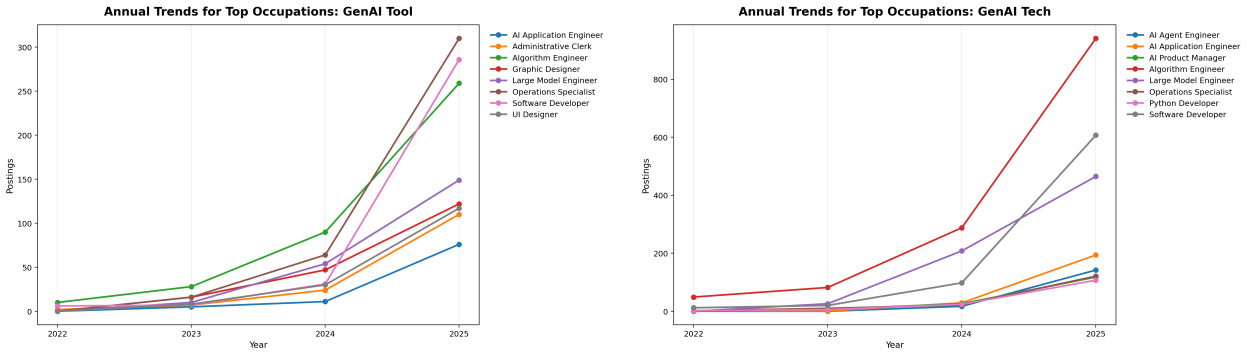


Figure 15: GenAI Adoption by Earlier AI Experience

Note: Firms are split by whether they posted at least one earlier-AI vacancy by 2022 ("AI adopters before 2022") or not ("No earlier AI"). Bars show the share of each group that has adopted GenAI (tool or tech) by the end of 2025, with listed firms in the left panel and non-listed firms in the right panel. Group sizes appear under each bar.

Fact 8: Adoption and potential job creation An analysis of job titles from postings requiring GenAI technical skills or tool knowledge shows rising demand for some GenAI-related occupations (Figure 16). Panel (a) shows that among postings requiring GenAI tool knowledge, the largest absolute increase in demand over the period 2022-2025 is for Operations Specialists, who manage and improve the day-to-day processes within an organization, followed by Algorithm engineers and software developers. Panel (b) further shows how the demand has increased for some occupations requiring GenAI technical skills. The largest increase in demand is among algorithm engineers, who develop algorithms for applications in machine learning, computer vision, and related fields. Comparing the two panels shows that the top occupations requiring GenAI technical skills are all ICT specialists, whereas those requiring GenAI tools also include non-ICT experts, such as UI Designer, Graphic Designer, and Administrative Clerk, who are highly exposed to GenAI but are not ICT workers (Gmyrek et al. 2025). This confirms that the impact of GenAI at the firm level is not confined to ICT workers or specialists.



(a) Annual trends: GenAI Tool

(b) Annual trends: GenAI Tech

Figure 16: Annual Trends in GenAI-Related Job Postings by Occupation

Note: Each line traces yearly postings from 2022 to 2025 for one of the top eight occupations in this sample. A steep upward slope indicates that the occupation became increasingly central to GenAI hiring over time.

We further analyze job titles in GenAI-related job postings to identify "new" occupations that may have emerged after the rise of GenAI. We rank the Top 15 such new occupations by skill requirements (i.e., GenAI technical skills vs tool knowledge) in Figure A4. Panel (b) shows that since 2022, the uptake of GenAI has created demand for engineers who specialize in GenAI technologies, while Panel (a) shows that it has also given rise to some non-ICT jobs, such as GenAI-related business development specialists. Both figures indicate that the need for proficiency in GenAI tools has spread well beyond technical specialists, whereas demand for greater GenAI technical skills remains largely concentrated among a narrow set of roles.

4.2 What Predicts GenAI Adoption?

The stylized facts listed above show that a range of firm characteristics are associated with GenAI adoption rates or intensity. Here, we estimate cross-sectional firm-level regressions to identify the firm characteristics most strongly associated with GenAI adoption by the end of 2025, conditional on the other included covariates.

Table 2 presents the estimation results. Columns (1) to (3) are linear probability models for whether a firm has ever adopted GenAI tools, GenAI techniques, or either. Column (4) applies the Poisson pseudo-maximum likelihood model to explain the adoption intensity (captured by the number of GenAI-related vacancies a firm posts). The correlates under consideration include: state ownership and log employment (both for listed firms only), broad-sector dummies (with manufacturing as the base), log total postings as a size measure, and a dummy for having posted

an earlier-AI vacancy by 2022. The two log terms are centered at their means, so the constant refers to a firm of average size. Panel A covers listed firms and Panel B the much larger pool of non-listed firms.

Table 2: Regression results: correlation between firm features and GenAI adoption

Dependent Variable	(1) GenAI Tool (=1) OLS	(2) GenAI Tech (=1) OLS	(3) GenAI (=1) OLS	(4) GenAI postings PPML
Panel A. Listed firms				
SOE	-0.915	-1.159	-0.931	0.153
	(1.110)	(1.177)	(1.248)	(0.212)
Market services	6.576***	10.593***	10.474***	0.705***
	(1.297)	(1.402)	(1.451)	(0.173)
Non-market services	3.690	-0.228	2.368	-0.403
	(2.645)	(2.671)	(2.904)	(0.288)
Construction, mining & utility	1.613	0.932	1.641	-0.675**
	(1.577)	(1.767)	(1.873)	(0.268)
Log total postings	4.018***	4.772***	5.614***	1.108***
	(0.242)	(0.265)	(0.286)	(0.090)
Log employment	3.890***	4.192***	4.203***	0.092
	(0.403)	(0.430)	(0.448)	(0.082)
Earlier-AI adopter by 2022	9.203***	13.361***	14.632***	0.845***
	(0.988)	(1.097)	(1.206)	(0.157)
Constant	8.204***	9.879***	12.510***	-1.906***
	(0.623)	(0.680)	(0.750)	(0.159)
Observations	5,183	5,183	5,183	5,183
R^2	0.162	0.208	0.223	–
Panel B. Non-listed firms				
Market services	-0.004	0.015**	0.003	0.185**
	(0.007)	(0.007)	(0.009)	(0.072)
Non-market services	-0.001	0.001	-0.008	-0.383***
	(0.008)	(0.008)	(0.010)	(0.116)
Construction, mining & utility	-0.029***	-0.016*	-0.046***	-0.310
	(0.008)	(0.009)	(0.011)	(0.254)
Log total postings	0.111***	0.143***	0.207***	0.984***
	(0.003)	(0.003)	(0.004)	(0.043)
Earlier-AI adopter by 2022	1.032***	2.368***	2.790***	1.465***
	(0.053)	(0.075)	(0.083)	(0.061)
Constant	0.078***	0.067***	0.129***	-7.203***
	(0.006)	(0.006)	(0.008)	(0.142)
Observations	2,438,447	2,438,447	2,438,447	2,438,447
R^2	0.005	0.013	0.014	–

Note: Columns (1)–(3) are linear probability models (OLS) for the adoption of GenAI tools, GenAI techniques, and either; column (4) is a Poisson pseudo-maximum-likelihood (PPML) model for the firm's total number of GenAI postings. Coefficients in columns (1)–(3) are in percentage points. Manufacturing is the omitted sector. Log total postings and log employment are centered at their sample means, so the constant is the predicted value for a baseline firm (a non-SOE manufacturer with no earlier AI) of average size. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

First, we find that firms with more postings are more likely to adopt GenAI and to do so intensively. This finding holds for both listed and non-listed firms. For listed firms alone, employment size is also positively and significantly associated with GenAI adoption and GenAI adoption intensity.

Second, sector matters for GenAI adoption as well, but mainly among listed firms. Listed market-service firms are about 10 percentage points more likely to adopt than manufacturers, while adoption probabilities for construction, mining, utilities, and non-market services differ little once size is controlled. In addition, market-service firms also post more GenAI vacancies than firms in other sectors.

Third, being an earlier AI adopter strongly predicts GenAI adoption after 2022. Listed firms that had posted an earlier-AI vacancy by 2022 are about 15 percentage points more likely to adopt GenAI by the end of 2025, and they post far more GenAI vacancies than those that had not adopted earlier AI. Similar effects are found for non-listed firms. Moreover, having adopted earlier AI technologies is also positively and significantly associated with the number of postings requiring GenAI tools/technical skills for both listed and non-listed firms.

5 Comparison with the U.S.

To put China's adoption patterns into context, we apply the same approach to job postings gathered from the United States. Specifically, we use job-posting data from Revelio Labs, a workforce analytics provider that compiles a large panel of online job postings and worker profiles from public sources and standardizes firms, job titles, occupations, industries, and locations. We apply the identical keyword dictionary to the U.S. postings, using the same GenAI tool, GenAI technique, and earlier-AI terms (Appendix Table A1) in English, to construct the same adoption measures.

Figure 17 presents the GenAI posting intensity since 2021 and the share of vacancies that mention a GenAI skill, split by tools and techniques. Among listed firms, the U.S. GenAI-related posting share reached about 1.5 percent (peaking near 2.4 percent in mid-2025) against roughly 0.8 percent in China by the end of 2025. By the end of 2025, non-listed firms in the U.S. surged to 1 percent of total postings, nearly four times the level for Chinese non-listed firms. Intensity jumps sharply in both countries around the release of DeepSeek, especially for non-listed firms in China, thereby sharpening competitive pressure in the U.S. In both countries, the technique share runs above the tool share, so firms hire not only to use off-the-shelf applications but also to build and adapt models.

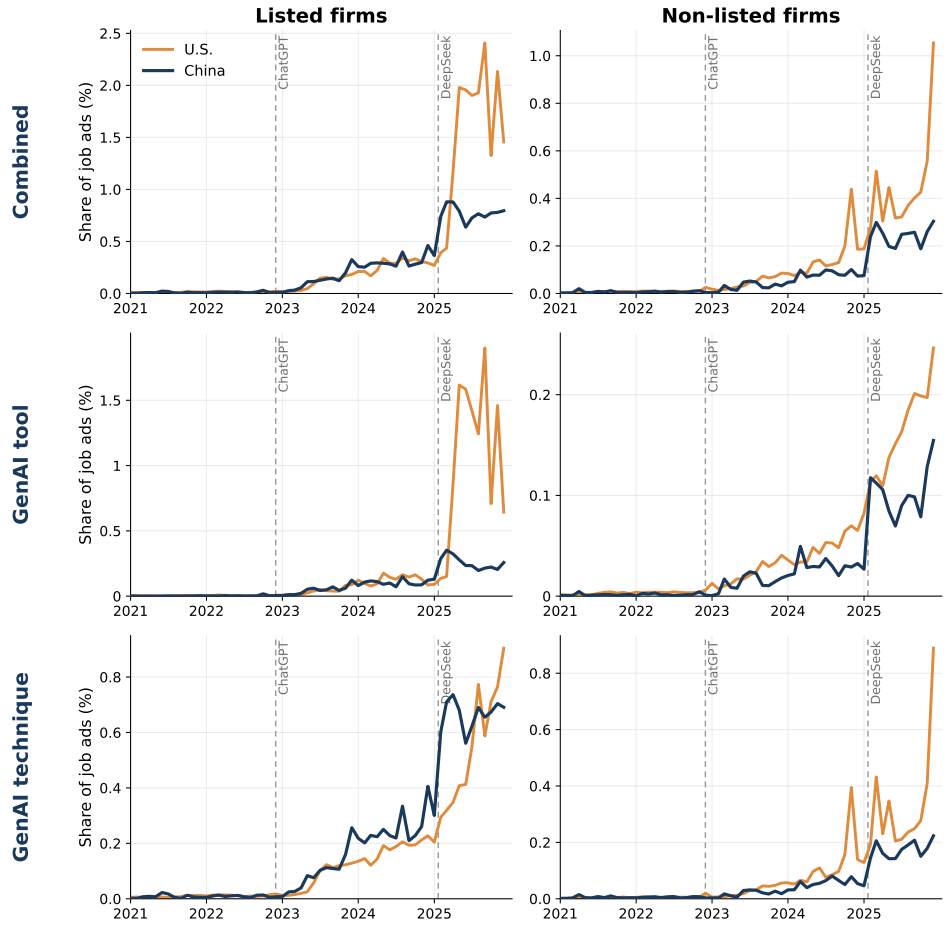


Figure 17: Monthly Share of Job Postings Mentioning a GenAI Keyword

Note: The figure shows posting-level GenAI intensity, the share of all job postings that mention a GenAI keyword, by month, January 2021 to December 2025. The dashed lines mark ChatGPT (November 2022) and DeepSeek (January 2025). The same keyword approach is applied to job postings in the U.S., sourced from Revelio Labs.

Which tools do firms hire for? Figure 18 classifies each GenAI-tool posting by the origin of the models: only U.S., only Chinese, both, or other. In 2024, the GenAI-tool postings of Chinese listed firms were dominated by U.S. models (i.e., 74 percent of identified postings only mentioned U.S. tools). By 2025, that share had fallen to 46 percent as domestic models surged, with postings naming Chinese models (alone or alongside U.S. models) rising from about a quarter to more than half among listed firms and from roughly a third to about half among non-listed firms. U.S. firms present a completely different picture: they hire almost entirely for U.S. models, with Chinese models accounting for about 0.1 percent for listed firms and around 1.5 percent for non-listed firms, even in 2025. The shift in model preferences in China coincides with the rise of competitive domestic models, including DeepSeek and Qwen, suggesting a rapid, partial substitution of

foreign tools by domestic tools.

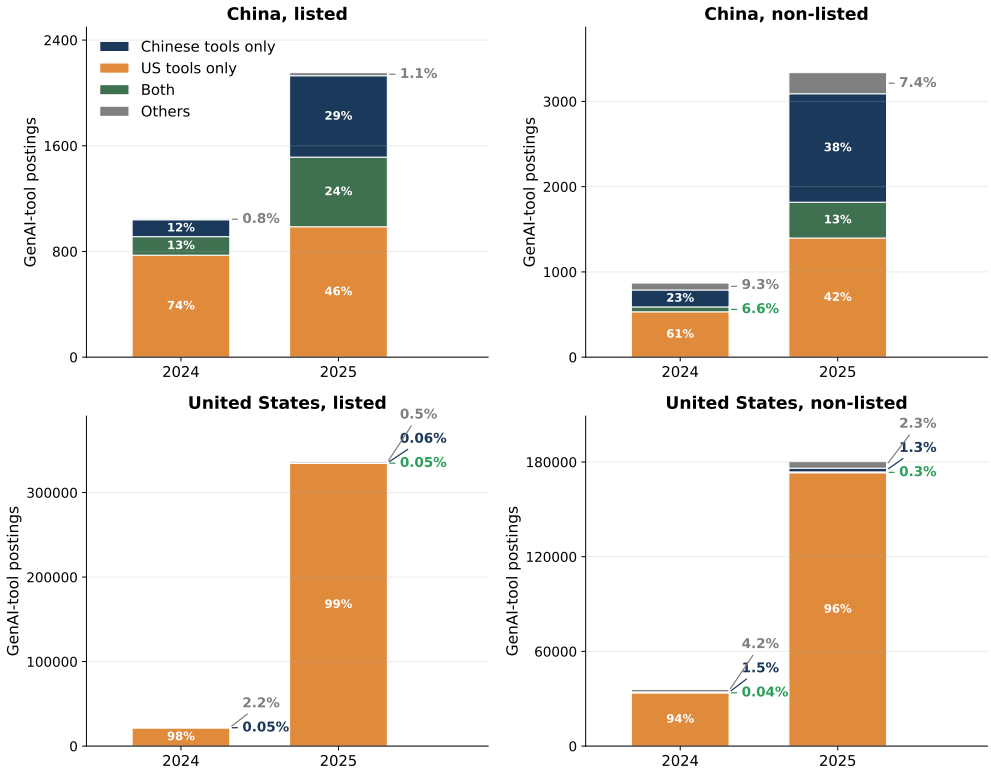


Figure 18: GenAI-Tool Postings by Tool Origin, 2024 versus 2025

Note: Each GenAI-tool posting is classified by the origin of the named model(s) it mentions: only US tools, only Chinese tools, both, or other. Origin is the headquarters country of the model's developer (e.g., ChatGPT, Copilot, Gemini, and Claude as U.S.; DeepSeek, Qwen, Doubao, Ernie, and Kimi as Chinese); "other" collects the few non-US, non-Chinese named tools, such as Mistral and Stable Diffusion. Bar height is the posting count, and labels give the within-bar share.

On the extensive margin, Figure 19 presents cumulative adoption at the firm level (i.e., the share of firms with at least one GenAI posting). Among listed firms, the U.S. leads the adoption race throughout. Noticeable GenAI hiring predates ChatGPT: by the time of ChatGPT's release, about 4 percent of U.S.-listed firms and roughly 1 percent of Chinese firms had already posted at least one job requiring a GenAI skill. By the end of 2025, 27 percent of the listed firms in the U.S. have done so, compared with 21 percent in China. Nevertheless, China's listed firms trail the U.S. frontier with a stable gap of about 4-6 percentage points between 2022 and 2025. In contrast, among non-listed firms, the adoption gaps widened between U.S. and Chinese firms. By the end of 2025, the adoption rate in the US is 1.5 percent, compared to 0.17 percent in China.

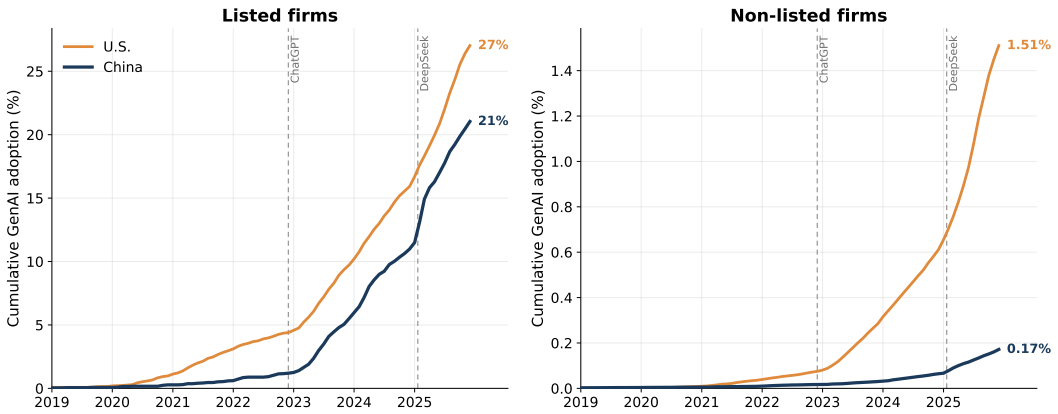


Figure 19: Cumulative GenAI Adoption Rate by Firms

Note: Each line traces the cumulative share of firms that have adopted GenAI, by month, from January 2019 to December 2025; a firm counts as an adopter from the month of its first job posting that requires a GenAI-related skill (tool or technique). The end-of-line labels give the end-2025 rate. The dashed lines mark the public release of ChatGPT (November 2022) and DeepSeek (January 2025). The U.S. series comes from Revelio Labs postings scored with the same keyword dictionary.

Lastly, Figure 20 compares the sector profile of listed-firm adoption. In China, adoption is concentrated in market services (35 percent) and declines across non-market services (23 percent), manufacturing (17 percent), and construction, mining, and utilities (15 percent). The adoption rate in the U.S. is higher overall and more evenly distributed across sectors: about 30 percent in market and non-market services and around 23-24 percent in manufacturing and the construction/mining/utility sectors. Market services is the only sector where the adoption rate in China exceeds that in the U.S. The results suggest that China's adoption is more concentrated, driven by services and technology firms, while the United States has a more broad-based pattern. Whether China's profile flattens toward the U.S. pattern as diffusion matures is a question these data raise but leave open.

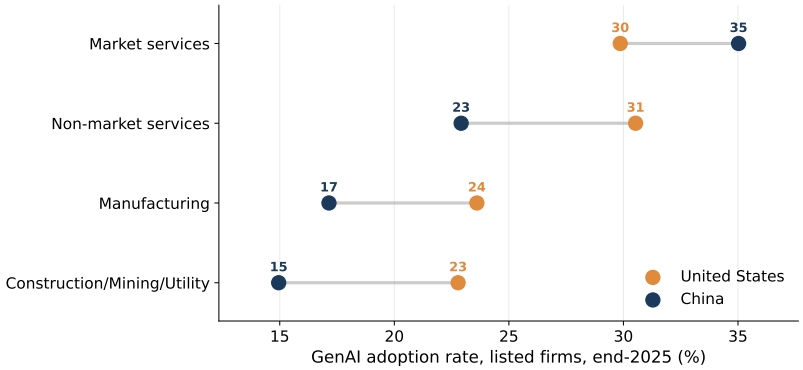


Figure 20: Listed Firms' GenAI Adoption Rate by Sector, End of 2025

Note: Each row shows the share of listed firms in a broad sector that had adopted GenAI (tool or technique) by the end of 2025. U.S. series comes from Revelio Labs postings scored with the same keyword dictionary.

6 Conclusion

This paper provides a job-posting-based account of how generative AI is diffusing across Chinese firms. The evidence points to a highly uneven pattern of diffusion. GenAI-related hiring has grown rapidly since the release of ChatGPT and accelerated further after the emergence of DeepSeek, but it remains concentrated among listed firms, larger firms, market-service firms, and firms with prior experience in earlier AI technologies. This pattern suggests that the diffusion of GenAI depends not only on access to low-cost tools but also on complementary organizational capabilities, technical readiness, and the ability to adjust hiring and internal functions. The comparison with the United States further shows that Chinese listed firms are approaching the adoption frontier, while the gap is much larger among non-listed firms.

The results suggest several directions for future research. First, combining job postings with firm surveys or administrative records would help distinguish between capability building and actual use. Second, linking GenAI adoption measures to firm performance, employment, and workforce composition would clarify the economic consequences of this new technology. Finally, the large gap between listed and non-listed firms highlights the importance of understanding what organizational and institutional frictions prevent smaller firms from translating accessible GenAI tools into productive adoption.

References

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

Alekseeva, Liudmila, José Azar, Mireia Giné, and Sampsa Samila (2025). "AI Adoption and the Demand for Managerial Expertise". Presentation at The Transformative Power of AI, European Central Bank, April 1, 2025.

Alekseeva, Liudmila, Mireia Giné, Sampsa Samila, and Bledi Taska (July 2020). "AI Adoption and Firm Performance: Management versus IT". Available at SSRN: https://ssrn.com/abstract=3677237.

Allen, Jeffrey S. (2026). "Monitoring AI Adoption in the U.S. Economy". *FEDS Notes*.

Babina, Tania (2026). *Understanding Firms' AI Efforts and Their Economic Impact*. NBER Working Paper 35123. National Bureau of Economic Research.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson (2024). "Artificial intelligence, firm growth, and product innovation". *Journal of Financial Economics* 151, p. 103745.

Babina, Tania, Anastassia Fedyk, Alex X He, and James Hodson (2023). *Firm investments in artificial intelligence technologies and changes in workforce composition*. Vol. 31325. National Bureau of Economic Research.

Bick, Alexander, Adam Blandin, and David J. Deming (2025). *The Rapid Adoption of Generative AI*. NBER Working Paper 32966. Originally issued September 2024; revised February 2025. National Bureau of Economic Research.

Bonney, Kathryn, Cory L. Breaux, Emin Dinlersoz, Lucia S. Foster, John C. Haltiwanger, and Aditya A. Pande (Apr. 2026). *The Microstructure of AI Diffusion: Evidence from Firms, Business Functions, and Worker Tasks*. NBER Working Paper 35141. National Bureau of Economic Research.

Bryson, Alex, John Forth, and Minghai Zhou (2014). "Same or Different? The CEO Labour Market in China's Public Listed Companies". *The Economic Journal* 124.574, F90–F108.

Calvino, Flavio, Lea Samek, Mariagrazia Squicciarini, and Cody Morris (2022). *Identifying and Characterising AI Adopters: A Novel Approach Based on Big Data*. OECD Science, Technology and Industry Working Papers 2022/06. Organisation for Economic Co-operation and Development.

Chen, Qin, Jinfeng Ge, Huaqing Xie, Xingcheng Xu, and Yanqing Yang (2025). "Large language models at work in China's labor market". *China Economic Review* 92, p. 102413.

Cheng, Zhaoqi, Dokyun Lee, and Prasanna Tambe (2023). "InnoVAE: Generative AI for Mapping Patents and Firm Innovation". *Available at SSRN 3482150*.

Cirera, Xavier, Diego Comin, and Marcio Cruz (2026). "Technology Sophistication Across Establishments". *The Quarterly Journal of Economics*.

Comin, Diego and Bart Hobijn (Jan. 2004). "Cross-Country Technology Adoption: Making the Theories Face the Facts". *Journal of Monetary Economics* 51.1, pp. 39–83.

Cong, Lin William, Xiaohan Yang, and Xiaobo Zhang (2024). "Small and Medium Enterprises Amidst the Pandemic and Reopening: Digital Edge and Transformation". *Management Science* 70.7, pp. 4564–4582.

Crane, Leland, Michael Green, and Paul Soto (Feb. 2025). "Measuring AI Uptake in the Workplace". *FEDS Notes*.

Dai, Min and Yukun Wang (2024). "Political connection and firm diversification: evidence from China". *Applied Economics* 56.50, pp. 6126–6143.

Dai, Ruochen and Xiaobo Zhang (2026). "E-commerce Expands the Bandwidth of Entrepreneurship". *The Journal of Law, Economics, and Organization*, pp. 1–28.

Eurostat (Dec. 2025). *Digital Economy and Society Statistics – Households and Individuals*. Statistics Explained Article. Data extracted December 2025. Eurostat, European Commission.

Fang, Hanming, Honglin Ren, Danwen Song, and Nianhang Xu (2023). *Environmentally-Inclined Politicians and Local Environmental Performance: Evidence from Publicly Listed Firms in China*. Working Paper 31071. National Bureau of Economic Research.

Gmyrek, Pawel, Janine Berg, Karol Kamiński, Filip Konopczyński, Agnieszka Ładna, Balint Nafradi, Konrad Rosłaniec, and Marek Troszyński (2025). *Generative AI and Jobs: A Refined Global Index of Occupational Exposure*. ILO Working Paper 140. Accessed: June 4, 2026. International Labour Organization.

Gmyrek, Pawel, Mariana Viollaz, and Hernan Winkler (2026). *Disruption without Dividend? How the Digital Divide and Task Differences Split GenAI's Global Impact*. Policy Research Working Paper 11328. Washington, DC: World Bank.

Goldfarb, Avi, Xianda He, and Florenta Teodoridis (2025). "Patterns of Artificial Intelligence Adoption by Hospitals". *AEA Papers and Proceedings* 115, pp. 40–45.

Hampole, Menaka, Dimitris Papanikolaou, Lawrence D. W. Schmidt, and Bryan Seegmiller (Feb. 2025). *Artificial Intelligence and the Labor Market*. NBER Working Paper 33509. National Bureau of Economic Research.

Hosseini, Seyed M. and Guy Lichtinger (Aug. 2025). "Generative AI as Seniority-Biased Technological Change: Evidence from U.S. Résumé and Job Posting Data". Preliminary manuscript, Harvard University.

Kalyani, Aakash, Nicholas Bloom, Marcela Carvalho, Tarek Hassan, Josh Lerner, and Ahmed Tahoun (May 2025). "The Diffusion of New Technologies". *The Quarterly Journal of Economics* 140.2, pp. 1299–1365.

Kureková, Lucia Mýtna, Miroslav Beblavý, and Anna Thum-Thysen (2015). "Using online vacancies and web surveys to analyse the labour market: A methodological inquiry". *IZA Journal of Labor Economics* 4, pp. 1–20.

Liu, Jessica and Douglas Webber (Mar. 2026). "AI Adoption and Firms' Job-Posting Behavior". *FEDS Notes*.

Liu, Yan, Jingyun Huang, and He Wang (2025). *Who on Earth Is Using Generative AI? Tracking Global Trends and Shifts in 2025*. Tech. rep.

Liu, Yan and He Wang (2026). "Who on Earth is using generative AI?" *World Development* 199, p. 107260.

McElheran, Kristina, J Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia Foster, and Nikolas Zolas (2024). "AI adoption in America: Who, what, and where". *Journal of Economics & Management Strategy*.

McElheran, Kristina, Mu-Jeung Yang, Zachary Kroff, and Erik Brynjolfsson (July 2025). "The Rise of Industrial AI in America: Microfoundations of the Productivity J-Curve(s)". Preliminary manuscript.

OECD (2025). *AI Adoption by Small and Medium-Sized Enterprises: OECD Discussion Paper for the G7*. Discussion Paper. Paris: OECD Publishing.

Shi, Hanwen, Anil K. Gupta, Waverly Ding, and Kunpeng Zhang (Nov. 2025). "Identifying AI Jobs in Online Job Postings: Existing Versus Newly Developed LLM-Based Approaches and A Dataset on Demand for AI Jobs among Russell 1000 Firms, 2018–2025". Working paper, Robert H. Smith School of Business, University of Maryland.

Song, Michael Zheng, Xiaohan Yang, and Xiaobo Zhang (2025). "AI Adoption Among Small- and Medium-sized Enterprises in China". Department of Economics, Chinese University of Hong Kong. Working paper.

Stone, Randall W., Yu Wang, and Shu Yu (2022). "Chinese Power and the State-Owned Enterprise". *International Organization* 76.1, pp. 229–250.

U.S. Census Bureau (Dec. 2025). *BTOS AI Core Question Updates*. Technical Note. U.S. Census Bureau.

Yao, Jiaquan, Kunpeng Zhang, Lipengc Guo, and Xu Feng (2024). "How Does Artificial Intelligence Improve Firm Productivity? Based on the Perspective of Labor Skill Structure Adjustment". *Management World (in Chinese)* 2, pp. 101–133.

Yotzov, Ivan, Jose Maria Barrero, Nicholas Bloom, Philip Bunn, Steven J. Davis, Kevin M. Foster, Aaron Jalca, Brent H. Meyer, Paul Mizen, Michael A. Navarrete, Pawel Smietanka, Gregory Thwaites, and Ben Zhe Wang (Feb. 2026). *Firm Data on AI*. Working Paper 34836. National Bureau of Economic Research.

Yuan, Zhe (2025). "Leveling the Playing Field: The Effect of AI on Small and Medium-Sized Firms". School of Economics, Zhejiang University. Working paper.

Zhang, Dandan, Hang Yu, Lixing Li, Jiayin Hu, Yiqing Mo, and Hongbo Li (2025). "The Measurement of AI Exposure and Its Impact on Labor Demand in China: Evidence from Large Language Models". *Management World (in Chinese)* 7, pp. 59–72.

Appendix

A Data

A.1 Details on the Firm-level Datasets

Listed firms In China, listed firms ("上市公司") are companies whose shares are publicly traded on the Shanghai or Shenzhen Stock Exchange. Under regulations administered by the CSRC, they are required to file audited annual reports ("年报") disclosing financial statements, ownership structure, board composition, employment level and structure, and other material information. This mandatory disclosure regime has made annual reports a central data source for empirical research on Chinese firms, such as Bryson, Forth, and Zhou (2014). The dataset covers about 5000 companies over the period 2005-2025, with the majority of firms from the manufacturing sector and larger in size than non-listed firms.

Non-listed firms. Firm-level information for non-listed firms is taken from China's official business registry ("中国企业工商注册信息"). All legally registered business entities in China are required to submit necessary information to the State Administration for Market Regulation (SAMR, formerly known as SAIC), thus forming the SAIC business registration database. The required information includes the unified social credit code, legal name, registered address, establishment date, registered capital, business scope, organizational form, and operational status. It also includes information on legal representatives, shareholders, and equity stakes, annual report filings, and any administrative penalties. By the end of 2024, the total number of registered business entities covered by this database stood at approximately 189 million, including 60.867 million enterprises. The recruitment data cover 2.8 million non-listed enterprises, accounting for approximately 4% of all enterprises.

The official business registry is publicly accessible through the National Enterprise Credit Information Publicity System (https://www.gsxt.gov.cn/index.html). More convenient access is provided through an online platform Qichacha ("企查查"), a leading commercial corporate information platform founded in 2014. Through a data partnership with the SAMR, the platform

provides real-time updates of business registration information.[37] For each business entity, the platform aggregates key registration details — including unified social credit codes, legal representatives, registered capital, and business scope (sector) — alongside shareholder and equity structures, subsidiary relationships, patent holdings, legal proceedings, and credit ratings. It has been widely used in academic research papers, such as Stone, Wang, and Yu (2022), Dai and Wang (2024), and Fang et al. (2023).

As non-listed firms do not report employment, we proxy firm size by the total number of job postings a firm makes over the sample period in Section 4. Figure A1 checks whether this proxy is valid by plotting the reported employment against the number of job postings for listed firms in our sample. Postings and employment line up well on a log scale, with a correlation of 0.42, supporting our usage of the number of postings as a proxy for firms' employment size.

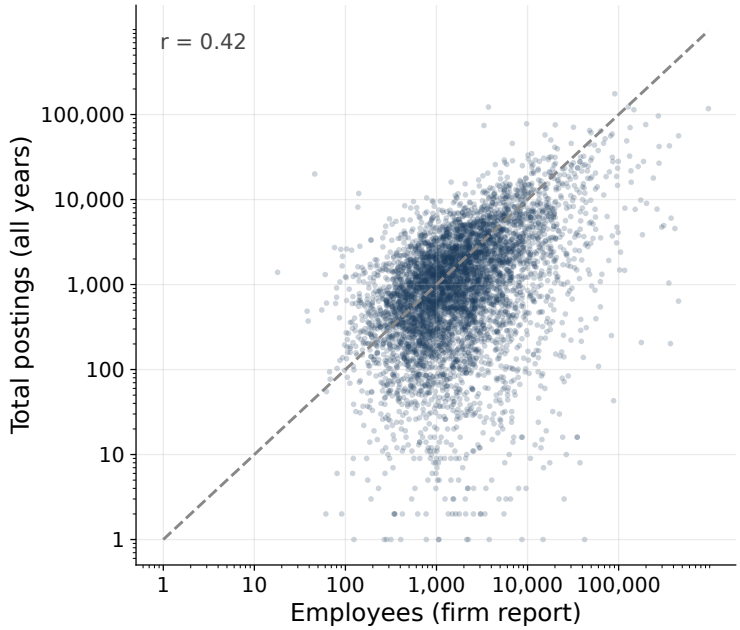


Figure A1: Total Postings versus Reported Employment, Listed Firms

Note: Each point is a listed firm. The horizontal axis reports employment, and the vertical axis is the total number of postings over the sample, both on a log scale. The dashed line is the 45-degree line. The two measures are positively correlated, with a log-log correlation of 0.42, which supports using total postings as a size proxy for firms without employment data.

[37] For more information, visit the company website: https://www.qcckyc.com/about-us-our-company.

A.2 Keyword Lists

The keyword list used to identify AI-related job postings is constructed around three conceptually distinct groups. The first group, GenAI Tool, captures specific generative AI products and platforms that firms may require employees to use, ranging from internationally prominent tools such as ChatGPT, Gemini, and Stable Diffusion to domestic Chinese platforms, including Doubao, Ernie Bot, Tongyi Qianwen, and Kimi chatbot. The second group, GenAI Tech, covers the technical methods, architectures, and workflows underlying generative AI development, such as prompt engineering, retrieval-augmented generation (RAG), fine-tuning, and large language models (LLMs). The third group, Earlier AI, encompasses broader machine learning and artificial intelligence concepts that predate the generative AI era, including deep learning, computer vision, natural language processing, and standard supervised learning methods.

Since the data derive from Chinese recruitment platforms, the keyword list incorporates both English-language terms and their commonly used Chinese equivalents. Many technical terms circulate in both languages within Chinese job postings, as practitioners frequently use the English acronym or brand name alongside the Chinese translation, so matching on either form is necessary to avoid undercounting. For keywords with no widely adopted Chinese equivalent, only the English form is matched; for Chinese-only expressions with no standard English counterpart, only the Chinese form is used. The full keyword list is reported in Table A1.

Table A1: AI Keyword List by Category

Category	Keyword List
GenAI Tool	ChatGPT, DeepSeek (深度求索), Gemini, Perplexity, Grok (格罗克), Claude, Copilot, Suno, Eleven Labs, Doubao (豆包), DeepAI, Poe, Kimi chatbot (Kimi智能助手), Blackbox AI, Tencent Yuanbao (腾讯元宝), Midjourney (MJ绘图), PixVerse, Meta AI, Leonardo, Ideogram, Zapier Chatbots, KREA, Prezi, Sora, NotebookLM, Le chat, v0.dev, Pixai, Nightcafe, OpusClip, Ernie Bot (文心一言), Youchat, ChatPDF, Synthesia, Pictory, Canva AI, Craiyon, Tongyi Qianwen (通义千问), Nano AI Search (Nano AI搜索), MaxAI, Playground, Easy-peasy, Writesonic, Liner, Murf AI, Stable Diffusion, DALL·E 3, Jasper AI, Phind, Artbreeder, Fliki, Luma Dream Machine, Designs.AI, Iflytek Spark (讯飞星火), Sparkdesk (星火大模型), Genmo AI, InVideo, 10Web AI Website Builder, Meta Image, StepFun (阶跃星辰), MiMo, GPT, GLM, Pangu Model (盘古大模型), Hunyuan Model (混元大模型), Zhipu AI (智谱), Baichuan (百川大模型), Minimax, Qwen (千问), 国产大模型, 自研大模型, Mistral, Falcon, 使用大语言模型, 使用生成式人工智能, 使用生成式AI, 使用代理式AI, 使用智能体AI
GenAI Tech	Fine-Tuning (模型微调), LangChain (LangChain框架), Prompt Engineering (提示工程), 提示词, Diffusion Model (扩散模型), Multimodal Model (多模态模型), AIGC (人工智能生成内容), 开发生成式人工智能, 开发生成式AI, Reinforcement Learning from Human Feedback (人类反馈强化学习), Reinforcement Learning from AI Feedback (AI反馈强化学习), PEFT (参数高效微调), CLIP (CLIP模型), RAG (检索增强生成), Retrieval-Augmented Generation, MultiModal Machine Learning (多模态机器学习), Diffusers (Diffusers库), LoRA (低秩适配), QLoRA, Multi-turn Dialogue (多轮对话), GAN (生成对抗网络), VAE (变分自编码器), Transformers (Transformer模型), LLM (大语言模型), Large Language Models, 开发大语言模型, AI-Generated Content (AI生成内容), 开发代理式AI, 开发智能体AI, Crowdsourcing Annotation (众包标注), Mixed Precision Training (混合精度训练), Hallucination (幻觉检测)
Earlier AI	Deep Learning (深度学习), Computer Vision (计算机视觉), Chatbot (聊天机器人), Machine Learning (机器学习), Machine Vision (机器视觉), Virtual Agents (虚拟智能体), Natural Language Processing (自然语言处理), NLP, Speech Recognition (语音识别), Pattern Recognition (模式识别), Object Recognition (目标识别), Neural Networks (神经网络), Supervised Learning (监督学习), Unsupervised Learning (无监督学习), Text Mining (文本挖掘), Mahout, Recommender Systems (推荐系统), Support Vector Machines (支持向量机), SVM, Random Forests (随机森林), Latent Semantic Analysis (潜在语义分析), Sentiment Analysis (情感分析), Opinion Mining (观点挖掘), Latent Dirichlet Allocation (潜在狄利克雷分配), Predictive Models (预测模型), Kernel Methods (核方法), Keras, Gradient Boosting (梯度提升), OpenCV, XGBoost, Libsvm, Word2Vec (词向量模型), Machine Translation (机器翻译), Sentiment Classification (情感分类), Decision Tree (决策树), Bayes (贝叶斯), Bayesian Methods (贝叶斯方法), 大数据处理框架, Large-scale Training (大规模训练), Distributed Training (分布式训练), Open-source Dataset (开源数据集), Parallel Training (并行训练), ImageNet

B Job Posting Deduplication Methodology

B.1 Overview

This appendix describes the deduplication pipeline applied to job posting data from major Chinese online recruitment platforms. Because firms routinely post identical or near-identical vacancies across platforms and refresh listings over time, deduplication is necessary to avoid inflating job counts in the analytical sample.

Listed and non-listed firms are treated differently owing to structural differences in their data sources. For **listed firms**, postings are drawn from three platforms (*51job, Zhilian, and Liepin*); cross-platform formatting variation precludes exact-match deduplication and necessitates a multi-stage fuzzy-matching pipeline. For **non-listed firms**, postings are sourced exclusively from *51job*; the uniform single-platform format permits a simpler, more stringent exact-match procedure. The key procedural differences are summarised in Table A2.

Table A2: *Comparison of Deduplication Procedures (Listed vs Non-listed Firms)*

Dimension	**Listed Firms**	**Non-listed Firms**
Data source	51job, Zhilian, Liepin (3 platforms)	51job only
Coverage period	2015–2025	2015–2025
Deduplication stages	3 stages: text preparation, strict, loose	2 stages: text cleaning, strict

B.2 Listed Firms

The data for listed firms are processed in the following three steps (Figure A2): (1) text preparation, (2) strict deduplication of exact-match postings, and (3) loose deduplication via location filtering and Jaccard similarity on job title and job responsibility.

B.2.1 Text Preparation

B.2.1.1 Extracting job responsibilities. To extract the sections on job responsibilities, we normalize the relevant text and implement a segment-based extraction approach on the job descriptions as follows:

- **Start-section detection.** Responsibility segments are identified by matching against a set of section-header keywords: 岗位职责, 工作内容, 主要职责, 工作职责, 主要任务, 职责, 职

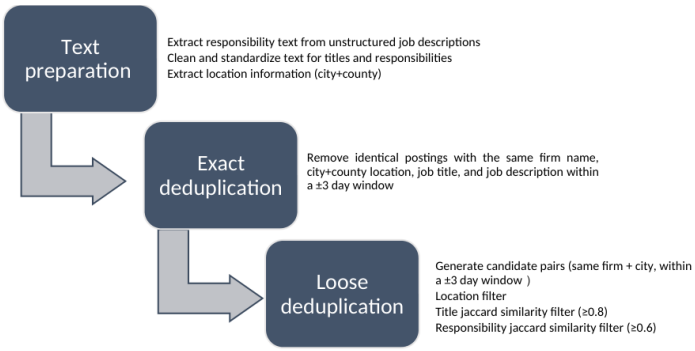


Figure A2: Flow chart of deduplication procedure – listed firms.

位描述, 职位说明, 岗位责任, 职位职责, 工作描述, 岗位描述, 岗位概述, tasks, duties, job responsibilities, job description, and responsibilities. Content is extracted from the earliest matched start keyword to the next section header or the end of the document.

- **End-section detection.** Segment boundaries are identified via a second keyword set: 任职资格, 任职要求, 职位要求, 岗位要求, 任职条件, 基本要求, 招聘要求, 岗位任职要求, 任职条件及要求, 应聘条件, 条件, 教育背景, 工作经验, 技能要求, 能力要求, 薪酬福利, 薪资待遇, 福利待遇, 我们提供, 待遇, 工作地点, 上班地址, 联系方式, 申请方式, 年龄, 工作时间, 职能要求, 公司简介, 公司介绍, 关于我们, requirements, qualifications.
- **Extraction hierarchy.** Multiple responsibility segments are concatenated. Text is extracted between both markers when available, before the end marker if only that is found, after the start marker if only that is found, and the full description if neither marker is present. Extractions shorter than 15 characters are flagged as failures. Extracted text is truncated to 300 characters to bound computational cost in subsequent similarity steps.

B.2.1.2 Text cleaning and standardisation. Job titles and extracted job responsibility text are cleaned through an identical procedure to remove platform-specific noise and improve matching accuracy. Specifically, we clean all text fields by converting strings to lowercase and applying Unicode NFC normalization, removing bracket delimiters while retaining their content (（）、【】、《》, (), []), stripping HTML tags, URLs, email addresses, and phone numbers, normalizing punctuation to spaces, removing all non-CJK, non-Latin, and non-digit characters, and collapsing whitespace. Job titles undergo additional cross-platform formatting as we remove recruitment

tags (e.g., 急聘、诚聘、五险一金, etc.), salary strings (e.g., "8k–12k"), city names, and trailing job ID codes (e.g., "J12345", "BJ-001").

B.2.1.3 Location variable construction. We need information on locations to identify cases where the same type of workers are hired for two positions in different locations. The raw address field follows a semi-structured format with "–" delimiters, showing province, city and country information in a sequence (e.g., "广东省 – 深圳市 – 南山区"). Here, we extract province, city, and county/district information to the extent possible. For regular cities, the district information is extracted from the third segment, which is empty if district information is unavailable.

B.2.2 Strict Deduplication (Listed Firms)

Exact-match duplicates are defined as postings from the same firm, at the same city-and-county location, with identical standardized titles and identical raw job descriptions, posted within overlapping time windows. This stage operates across platforms, capturing within-firm cross-platform reposts. Time proximity is operationalized by assigning each posting to a 7-day bin and expanding it by ± 1 bin to create overlapping weekly windows. Matching is performed on the *raw* (uncleaned) description to ensure only character-for-character identical postings are collapsed.

B.2.3 Loose Deduplication (Listed Firms)

Strict deduplication removes character-for-character identical postings but misses near-duplicates arising from platform-specific formatting differences or minor firm-side adjustments across sites—such as slight title variations (e.g., "软件工程师" vs."软件开发工程师") or description rewording and appended boilerplate. Loose deduplication addresses this via a Jaccard similarity pipeline applied to cleaned text fields, with an early location filter to improve computational efficiency and matching precision:

- **Candidate pair generation.** We block on firm name and city within a ± 3-day posting window and form candidate pairs via a three-way join on firm name, city, and time bin. Pairs with an actual date gap exceeding three days and self-pairs are excluded.

- **Location filter.** Before computing text similarity, we drop pairs for which both postings carry county information and the counties differ, as postings across counties within the same city typically reflect distinct office locations.[38]
- **Title similarity.** For each surviving pair, we compute the Jaccard index over character-level bigrams of the standardized titles,

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|},$$

  where A and B are the bigram sets. We discard pairs with $J < 0.8$.
- **Responsibility similarity.** Pairs passing the title filter are evaluated on cleaned responsibility text using the same character-bigram Jaccard measure. Only pairs with $J \geq 0.6$ are retained as duplicates.
- **Clustering and representative selection.** Duplicate pairs may form transitive chains: if $a \sim b$ and $b \sim c$, then $\{a, b, c\}$ form one cluster. We resolve clusters via iterative label propagation, initializing each posting with its own ID and updating each label to the minimum among its neighbors until convergence, capping at 50 iterations.

B.3 Non-Listed Firms Pipeline

Here, we use a two-step deduplication procedure for non-listed firms because their postings originate exclusively from 51job, making exact-match deduplication both feasible and sufficient.[39] First, both job title and job description are cleaned via a unified procedure: strings are converted to lowercase and Unicode NFC normalized; bracketed content is removed across all bracket types (（…）、【…】、《…》, (…), […]); HTML tags, URLs, email addresses, and phone numbers are stripped; all punctuation—ASCII and Chinese full-width—and digits are removed; and any remaining non-CJK, non-Latin characters are dropped before collapsing and trimming whitespace. Then, strict deduplication is applied where duplicates within a 7-day overlapping time window are dropped (conditioning on firm name, location, title, time bin, and cleaned description).

[38] County information is missing for some listed-firm postings; we therefore retain any pair for which county data are absent in either posting.

[39] As a robustness check, the full three-stage listed-firms pipeline was applied to non-listed firms. The resulting loose-duplication removal rate was negligible, confirming that the simplified procedure introduces no material bias.

C Classify AI-related Job Postings by Business Function

To better understand AI adoption across business functions within firms, we develop a function classification system based on job postings. We define five functional categories — technical, production, sales, administrative, and financial — and implement a three-tier classification pipeline (Figure A3) to assign each posting to its corresponding function.

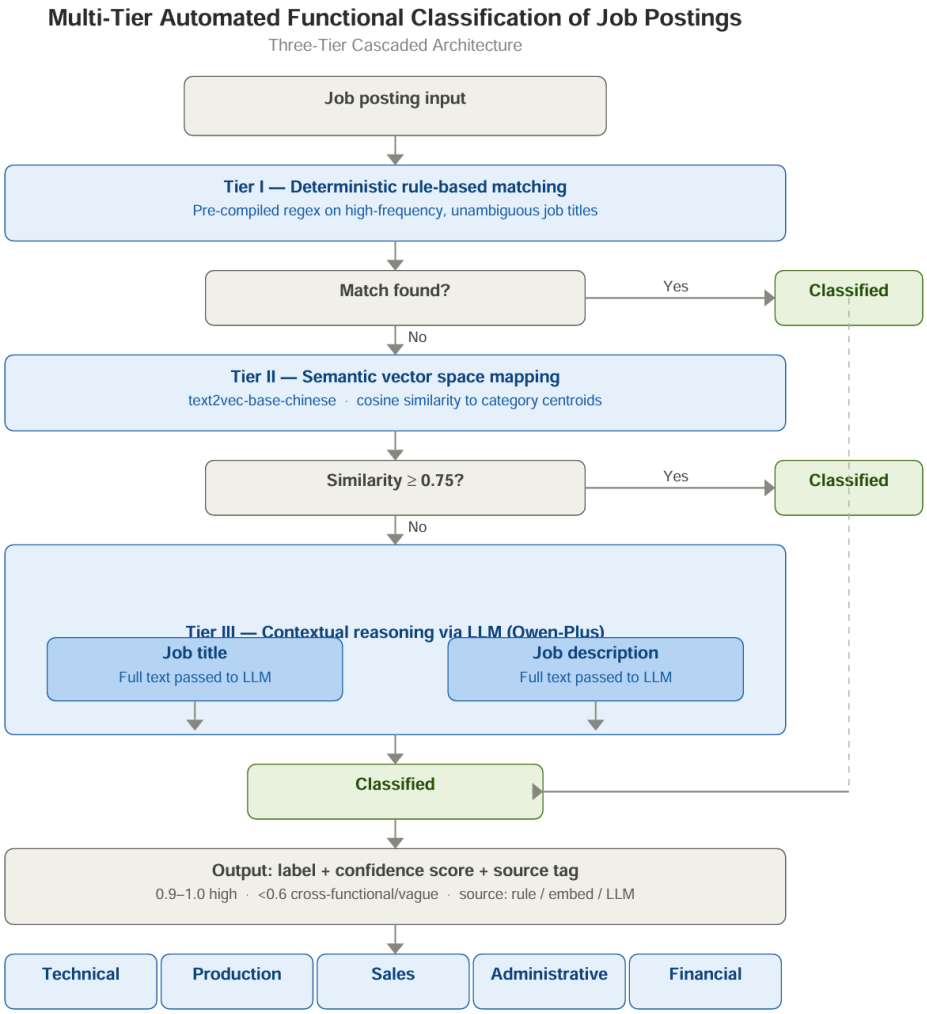


Figure A3: Three-tier classification pipeline for job postings.

C.1 Functional Taxonomy

Job postings are classified into five mutually exclusive business functions based on their core organizational role:

- **Technical (R&D):** Roles focused on software development, algorithm design, hardware research, and technical architecture, including algorithm engineers, data scientists, and chief technology officers.
- **Production (Product & Delivery):** Roles centered on product management, project delivery, manufacturing processes, and data operations, including product managers, DevOps engineers, and data labeling specialists.
- **Sales (Commercial & Growth):** Roles oriented toward revenue generation and market expansion, including business development, sales, marketing, and customer success.
- **Administrative (Support & Compliance):** Roles providing organizational support, including human resources, legal, compliance, and executive administration.
- **Financial (Finance & Risk):** Roles focused on capital management, accounting, and risk monitoring, including auditors, investment analysts, and quantitative traders.

C.2 Three-Tier Classification Pipeline

To accommodate the inherent ambiguity in job titles and descriptions, classification proceeds through a cascaded three-tier architecture, combining rule-based matching, semantic embeddings, and large language model (LLM) reasoning.

Tier I: Rule-Based Matching. Tier I applies deterministic keyword matching to job titles. For each of the five functional categories, we compile a list of high-frequency, unambiguous title keywords drawn from domain knowledge of Chinese labor markets (Table A3). A posting is classified at this tier if its title contains an exact match to any keyword in the corresponding list. For instance, titles containing terms such as 销售" or 市场" are assigned to the Sales category, while titles containing 算法" or 工程师" are assigned to Technical. The keyword lists are implemented as pre-compiled regular expressions, enabling efficient batch processing of large-scale job posting datasets. Since this tier relies solely on explicit title signals, it achieves high precision for standardized roles and serves as the primary classification layer, handling the majority of postings before any model-based inference is required.

Table A3: Functional categories and representative keywords (Tier I)

Category	Representative Keywords
Sales	销售, 商务, 业务员, 市场, 品牌, 营销, 售前, 售后, 客户经理, 招商, marketing, sales
Financial	财务, 会计, 出纳, 成本, 结算, 报税, 审计, 投资, 量化, 风控, 税务, finance, fintech
Administrative	行政, 文员, 前台, 运营, 人事, 招聘, 法务, 合规, 秘书, HR, admin, legal
Production	产品经理, 产品助理, 产品专员, 项目经理, 交付, 制造, 数据标注, 信息录入, 数据录入, 内容审核, 标注审核, 数据采集, AI训练师, 操作工, 普工, 装配, 质检, 检验, 车间, 生产工, 生产计划, 仓库, 物料, 包装, 机台, production
Technical	算法, 开发, 研发, 工程师, 架构, 程序员, 后端, 前端, 测试, 模型, 大数据, 数据分析, 数据科学, 技术, 接口, NLP, engineer, developer, algorithm, DevOps, SRE, SaaS, PaaS

Tier II: Semantic Embedding Similarity. Titles that do not match any Tier I keyword are passed to Tier II, which applies semantic vector similarity to capture roles that share functional meaning with the keyword list but do not contain an exact lexical match. We employ the `text2vec-base-chinese` sentence embedding model, a transformer-based model pre-trained on large-scale Chinese corpora that encodes semantic relationships between words and phrases in Chinese-language text. For each functional category, we construct a category centroid by averaging the embeddings of its representative keywords (Table A3). We then compute the cosine similarity between the embedding of each unmatched job title and each of the five category centroids. A posting is assigned to the category with the highest similarity score, provided that score meets or exceeds a threshold of 0.75. Postings whose maximum similarity falls below this threshold are considered ambiguous and forwarded to Tier III for further reasoning.

Tier III: LLM Contextual Reasoning. Postings that fall below the Tier II similarity threshold are forwarded to Tier III, which employs a large language model (LLM) to perform contextual reasoning over both the job title and the full job description (JD). Unlike the preceding tiers, which rely on lexical or distributional similarity to a fixed keyword list, Tier III is designed to handle genuinely ambiguous roles where functional identity cannot be inferred from the title alone. We

use Alibaba's Qwen-Plus model, prompted with a structured expert system that encodes category definitions, boundary rules, and disambiguation logic for high-frequency ambiguous terms, as shown in Listing 1.

The prompt instructs the model to prioritize functional content over seniority level, so that senior roles such as directors and vice presidents are classified by their business domain rather than their rank. It also supplies explicit branching rules for commonly ambiguous terms. For instance, a procurement role is assigned to Production if its focus is raw materials or supply chain management, to Administrative if it concerns office supplies or indirect sourcing, and to Financial if the primary responsibility is budget control or auditing. Similarly, operations roles are routed to Technical for infrastructure or DevOps work requiring programming, and to Administrative for office or facilities management. For English-language postings, the model is instructed to translate job responsibilities before classification to avoid misleading surface-level terms such as "production environment."

The model returns a Python-formatted list containing the assigned category and a confidence score between 0 and 1. Scores of 0.9 or above indicate unambiguous postings with explicit functional signals; scores below 0.6 flag cross-functional roles or vague job descriptions for potential manual review. In cases where responsibilities span two clearly distinct functions with no discernible primary role, the model may assign up to two categories.

Listing 1: Tier III classification prompt (Qwen-Plus)

```
You are an expert HR job function classifier. Based on the job title
and description, assign the posting to one of five functional categories.

### Category definitions:
1. Technical: Core focus on coding, algorithms, hardware R&D, or
   technical architecture (e.g., algorithm engineers, data scientists,
   DevOps, CTOs).
2. Production: Core focus on product definition, project delivery,
   manufacturing processes, or data processing (e.g., product managers,
   AI trainers, data labeling, industrial design).
3. Sales: Core focus on revenue generation and market expansion
   (e.g., BD, sales, marketing, customer success).
4. Administrative: Core focus on organizational support, legal, and
   HR compliance (e.g., HR, admin, legal, compliance, secretaries).
5. Financial: Core focus on capital management, accounting, and risk
   monitoring (e.g., auditors, quant traders, investment analysts).

### Disambiguation rules for ambiguous terms:
- Procurement: -> Production if raw materials/supply chain;
               -> Administrative if office supplies/indirect sourcing;
               -> Financial if budget control/auditing.
- Operations:  -> Technical if DevOps/SRE/coding required;
               -> Production if IT application maintenance/data ops;
               -> Administrative if office/facilities management.
- Business/BD: -> Sales if client acquisition with revenue targets;
               -> Production if supply-chain negotiation/contract drafting.

### Classification rules:
1. Output only the single most relevant category by default.
2. Output two categories only if responsibilities clearly span two
   domains with no discernible primary function.
3. Prioritize functional content over seniority level.
4. For English postings, translate responsibilities before classifying.
5. Assign a confidence score from 0.0 to 1.0:
   - 0.9-1.0: Information is explicit (e.g., Java software engineer).
   - 0.7-0.8: Category is clear but JD is somewhat vague.
   - 0.5-0.6: Cross-functional role or highly ambiguous description.

Job title: {p_title}
Job description: {p_jd}

Output format (Python list only, no explanation):
[["Category"], confidence]
Examples: [["Technical"], 0.98] / [["Production", "Sales"], 0.75]
```

D Additional Tables and Figures

Table A4: Job Posting–Listed Firms Sample Matching Rate

Year	Matched sample	Full listed firms sample	Matched rate
2015	3,616	4,159	86.9%
2016	3,959	4,406	89.9%
2017	4,270	4,721	90.4%
2018	4,562	5,006	91.1%
2019	4,738	5,151	92.0%
2020	4,770	5,181	92.1%
2021	4,741	5,183	91.5%
2022	4,787	5,183	92.4%
2023	4,742	5,183	91.5%
2024	4,672	5,182	90.2%
2025	4,943	5,182	95.4%

Note: The listed firms database is sourced from WIND.

Table A5: Number of Job Postings (millions) Before and After Deduplication, Listed Firms

Year	After Deduplication	Before Deduplication	Duplicates Removed
2015	0.93	1.02	0.09
2016	1.16	1.29	0.14
2017	1.58	1.77	0.19
2018	1.69	1.90	0.21
2019	1.60	1.85	0.25
2020	1.10	1.29	0.19
2021	1.57	1.83	0.26
2022	1.76	2.06	0.31
2023	1.41	1.63	0.23
2024	1.03	1.20	0.17
2025	0.86	1.01	0.15
Total	14.68	16.85	2.18

Notes: Figures in millions of job postings. Data of listed firms' job postings are sourced from three major Chinese recruitment platforms (51job, Zhaopin, Liepin), 2015–2025. Duplicates are identified via a three-stage pipeline as indicated in Appendix Section B.

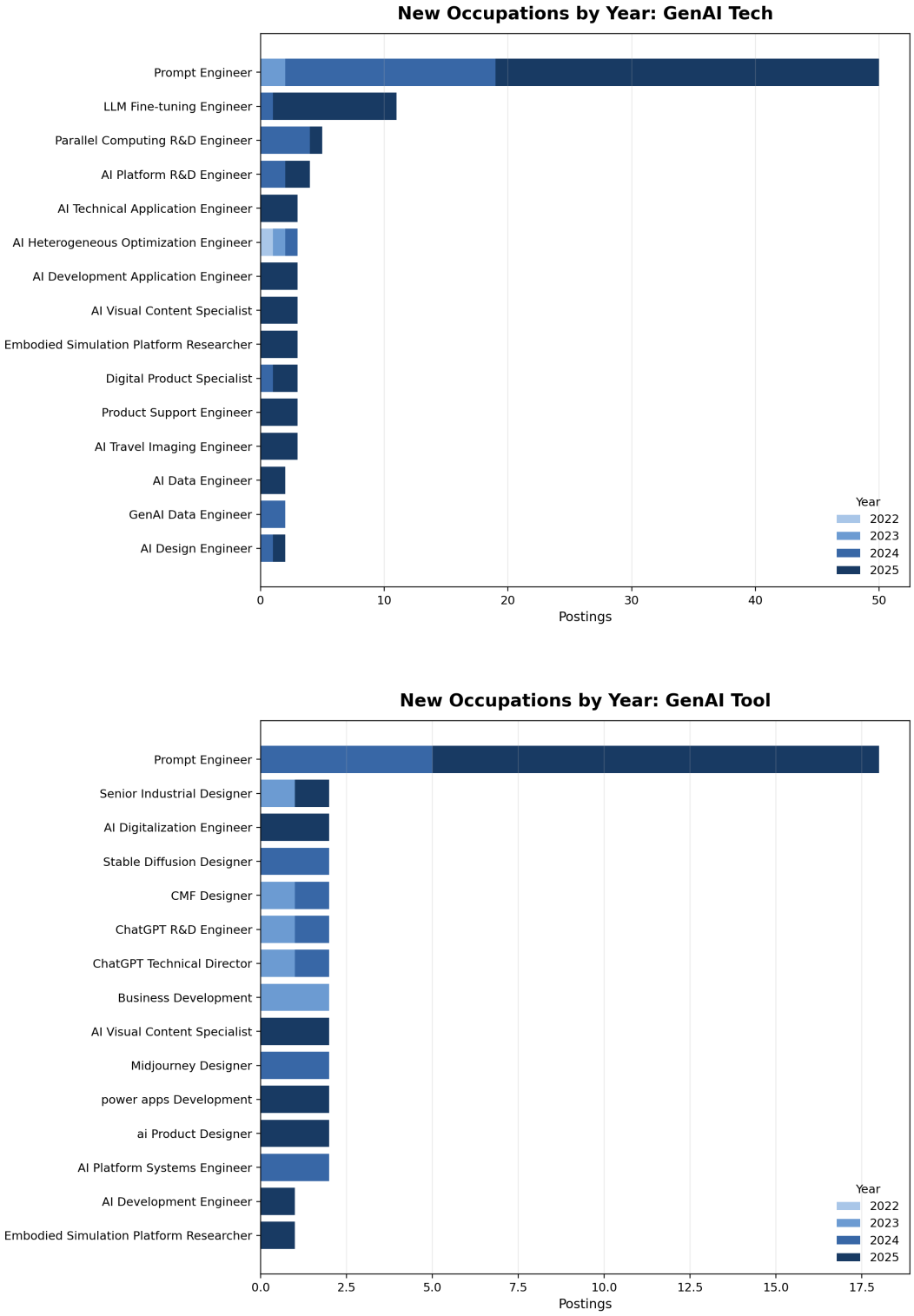


Figure A4: New GenAI-related Occupations

Note: The stacked bars show occupations that did not appear before 2022 in the full posting data but appear in post-2022 GenAI hiring. Colors decompose each occupation's total into the first observed and subsequent years.

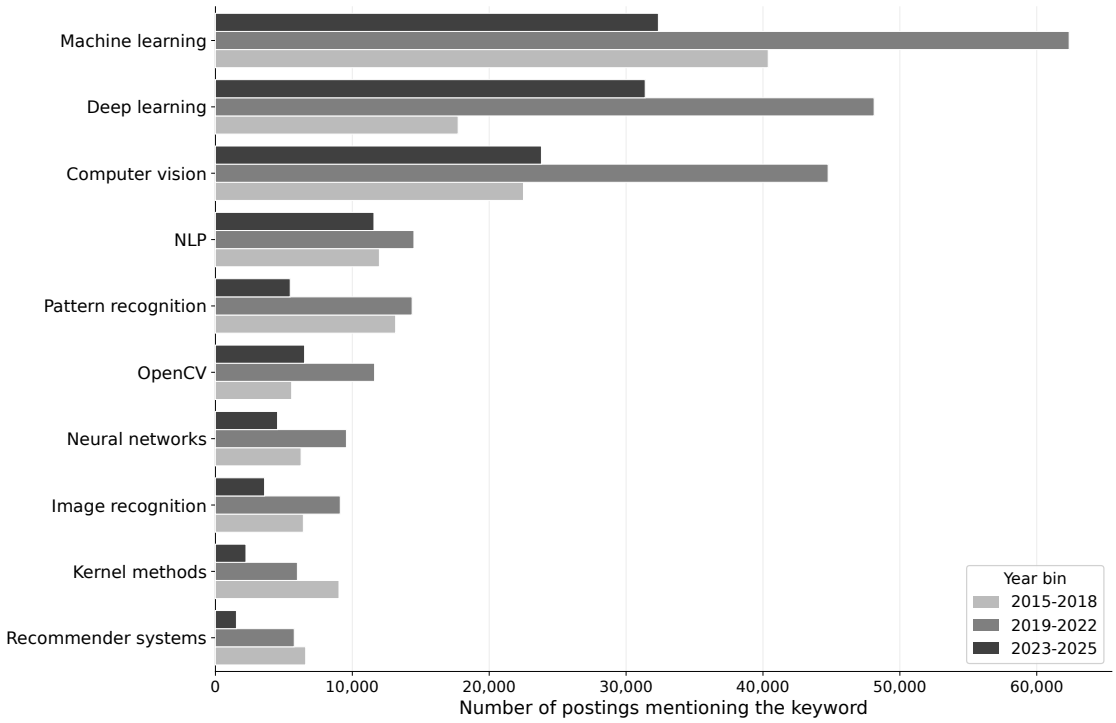


Figure A5: Top Earlier AI Keywords

Note: The ten most frequent earlier-AI keyword concepts by number of postings, grouped into 2015–2018, 2019–2022, and 2023–2025.

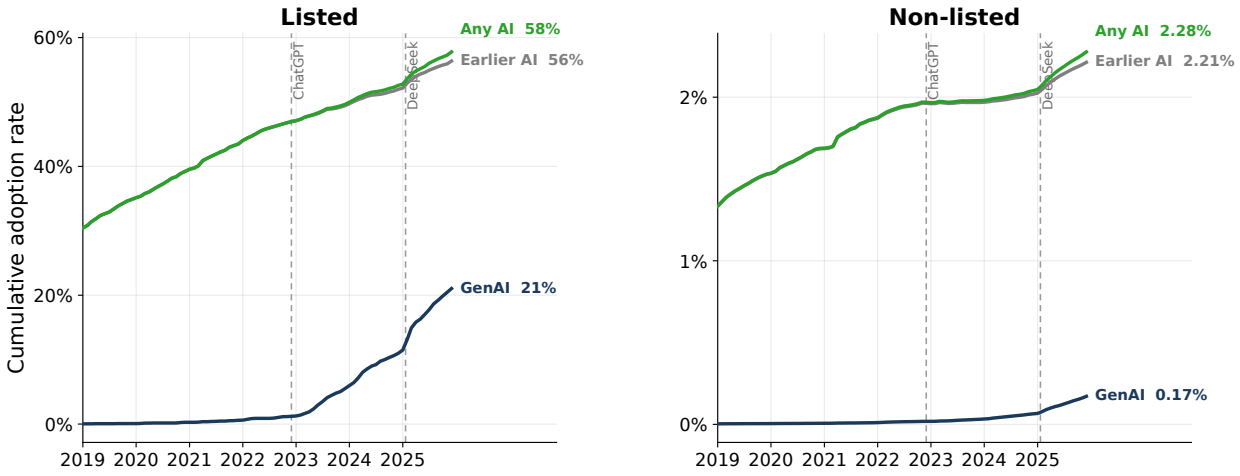


Figure A6: Cumulative Adoption Rate

Note: Monthly cumulative share of firms that have adopted earlier AI, GenAI, or any AI. By the end of 2025, among listed firms, the rates are 58 percent (any AI), 56 percent (earlier AI), and 21 percent (GenAI); among non-listed firms, they are 2.3, 2.2, and 0.2 percent. Counterpart of Figure 5.

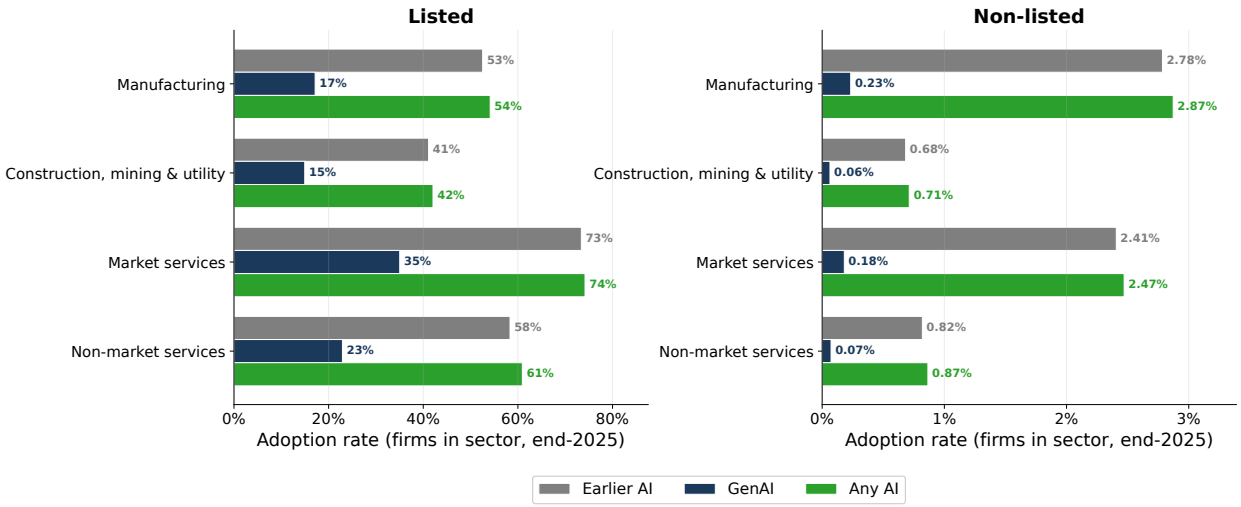


Figure A7: Adoption by Broad Sector, End of 2025

Note: Adoption rate by broad sector for each measure, end of 2025, for listed and non-listed firms. Counterpart of Figure 7.

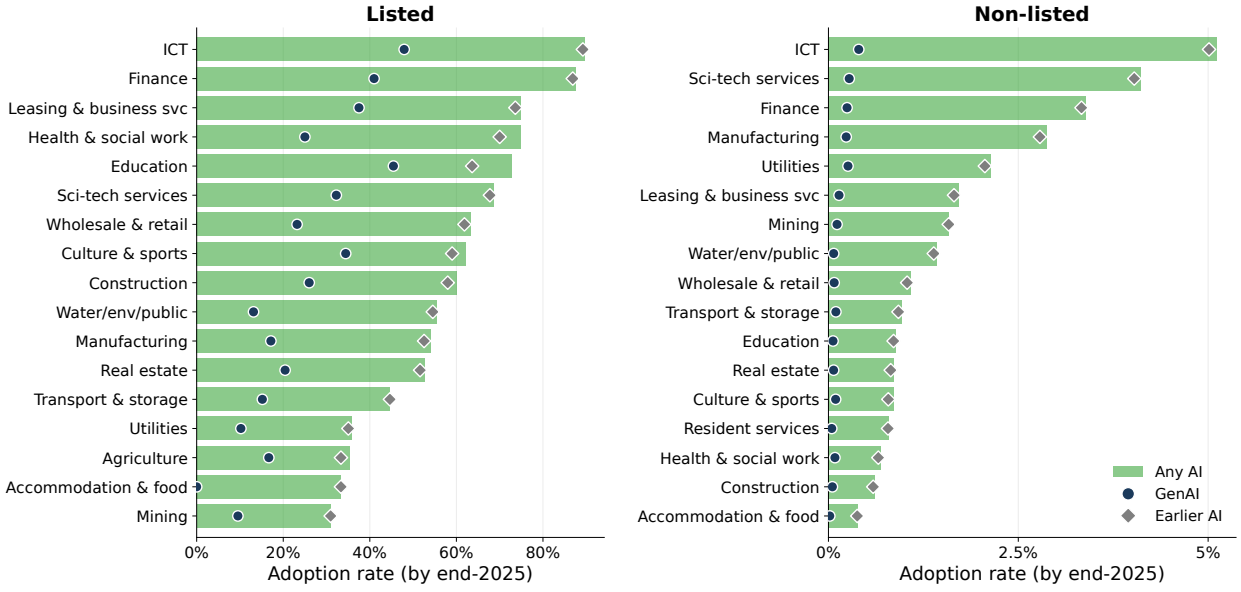


Figure A8: Adoption by Industry, End of 2025

Note: Adoption rate by GB/T industry section for each measure, end of 2025. Information and communication technology leads on every measure. Panels use independent x-scales. Counterpart of Figure 8.

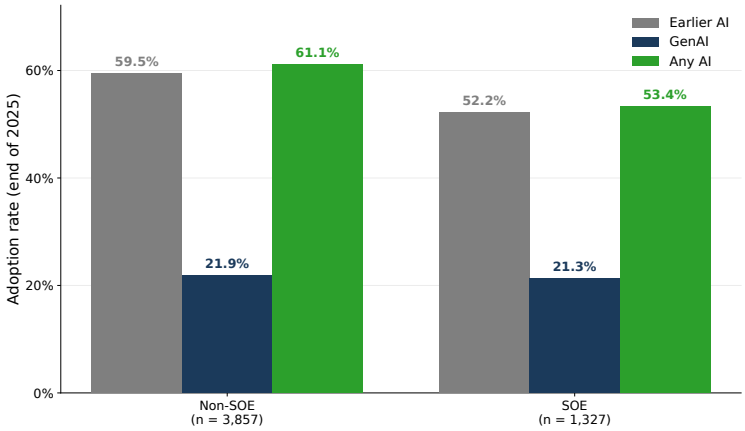


Figure A9: Adoption by Ownership, by Measure (Listed Firms, End of 2025)

Note: Adoption rate among listed firms by ownership type for each measure, end of 2025. Non-SOEs adopt at higher rates than SOEs on every measure. Counterpart of Figure 11.

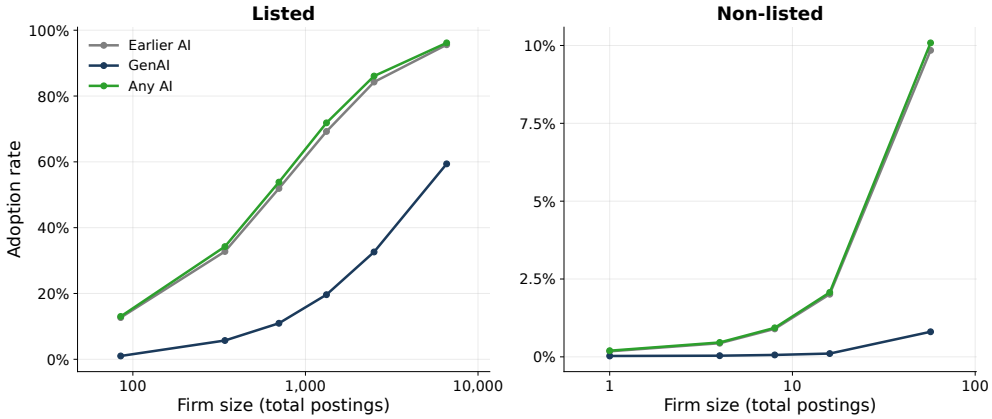


Figure A10: Adoption by Firm Size, by Measure

Note: Adoption rate by firm-size bin (total postings, six equal-count bins) for each measure. Adoption rises with size on every measure. Panels use independent scales. Counterpart of Figure 10.