Technology Diffusion in Emerging Markets: Evidence from China*

Danqing Mei† Hongshen Sun‡ Tianshu Sun§

September 24, 2025

ABSTRACT

We study how new technologies diffuse in an emerging-market setting, using China as a case. Linking 31 million job postings (2016–2021) with Chinese patents (2000–2022), we track 3,269 technologies from invention into labor demand. As in advanced economies, technologies emerge in a few hubs and gradually broaden across regions. But the pace of diffusion is shaped by emerging-market frictions. Limited pools of high-skill labor slow the spread of research-intensive and foundational technologies, while application-oriented jobs diffuse more quickly. Reliance on foreign knowledge creates a dependence drag, as U.S.-rooted technologies diffuse more slowly. External shocks also redirect adoption: COVID-19 slowed diffusion, especially in in-person occupations, while U.S. sanctions accelerated it through substitution and localization. These results show how skills, complexity, and global interdependence condition diffusion in emerging economies.

*This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors declare no conflicts of interest.

†Cheung Kong Graduate School of Business. He can be reached at dqmei@ckgsb.edu.cn.

‡MIT Sloan School of Management. He can be reached at sunhs@mit.edu.

§Cheung Kong Graduate School of Business. He can be reached at tianshusun@ckgsb.edu.cn.

I. Introduction

Why does the pace of technology diffusion differ so markedly across economies? A large literature has documented how new technologies emerge in a few hubs in advanced economies and then spread only slowly across space and occupations (e.g., Griliches, 1957; Jaffe et al., 1993; Comin and Hobijn, 2004; Keller, 2004; Kalyani et al., 2025). Much less is known about diffusion in emerging markets, where labor markets are thinner, institutions differ, and reliance on foreign knowledge and inputs is greater. These features create distinctive frictions that condition how innovation translates into broad-based economic opportunity. Understanding these frictions is central to evaluating the growth prospects of emerging economies and to designing policies that support catch-up.

This paper studies *emerging-market frictions in technology diffusion*, using China as a case. Once regarded primarily as a technology follower, China has become a leading source of patents and innovation output. Hu and Jefferson (2009) document the surge in Chinese patenting, attributing much of it to policy incentives rather than immediate gains in quality. More recently, König et al. (2022) show that Chinese firms have transitioned from imitation toward genuine innovation with measurable productivity effects. At the same time, China's innovation trajectory is tightly linked to its relationship with the United States. Han et al. (2024) show how U.S.–China technological interdependence, measured through patent citations, has been reshaped by policies of decoupling, with direct implications for firm performance. These developments highlight China as a salient example of how emerging-market conditions—thin skill pools, dependence on foreign technology, and vulnerability to shocks—shape diffusion.

We construct a new dataset that links more than 31 million Chinese job postings (2016–2021) with all patents granted in China between 2000 and 2022. Using large language models to classify occupations and skills, and patent text to identify technologies, we isolate 3,269 technologies that are both innovative (patented) and relevant to labor demand (appearing in postings). This framework allows us to track how technologies move from invention into

adoption across Chinese cities and occupations.

Our analysis yields several findings. Technologies in China broaden across regions over time, consistent with patterns documented in advanced economies. The pace of diffusion, however, is systematically shaped by emerging-market frictions. Technologies tied to low-skill, application-oriented jobs spread more rapidly than those requiring high-skill, research-intensive labor, reflecting the uneven distribution of talent. Foundational or upstream technologies diffuse more slowly than downstream applications, and technologies more dependent on U.S. prior art diffuse more slowly than those rooted in domestic knowledge. Finally, external shocks matter: the COVID-19 pandemic slowed diffusion—especially for occupations requiring physical presence—while U.S. Entity List sanctions accelerated diffusion, consistent with substitution and localization.

These results show that diffusion is mediated not only by market incentives but also by the interaction of labor supply, technology position, and international exposure. In the Chinese case, the geography of adoption reflects the limited depth of high-skill labor pools, the degree of dependence on foreign knowledge bases, and the impact of shocks that redirect diffusion trajectories.

Our paper contributes to several literatures. We build on classic studies of technology adoption and diffusion (Griliches, 1957; Comin and Hobijn, 2004; Keller, 2004), the geography of knowledge spillovers (Jaffe et al., 1993; Hall et al., 2001), and recent work on the diffusion of new technologies (Kalyani et al., 2025). Relative to Kalyani et al. (2025), who analyze the United States, we complement their findings by showing how diffusion operates in an emerging-market setting, where skill constraints, foreign dependence, and shocks are first-order. Methodologically, our approach is in the spirit of Hershbein and Kahn (2018) in using vacancy data to study technology adoption, but differs in assembling and classifying an entirely new dataset of over 30 million postings in China. Substantively, we connect to work on regional inequality and place-based policy (Moretti, 2012; Kline and Moretti, 2014) and to research on China's innovation trajectory and U.S.–China technological linkages (Hu

and Jefferson, 2009; König et al., 2022; Han et al., 2024).

The most closely related study is Kalyani et al. (2025), who document diffusion dynamics in the United States using patents and job postings. Their results establish baseline stylized facts for a mature, market-driven innovation system: technologies originate in high-skill regions and spread slowly, with pioneer hubs retaining an advantage in research and development. Our paper complements this evidence by showing how diffusion unfolds in China, an emerging economy with a state-led innovation model and deep global interdependence. By analyzing both regional and skill heterogeneity and the role of external shocks, we highlight how emerging-market frictions condition the speed and breadth of diffusion. Taken together, the two papers provide a comparative perspective on how technologies diffuse in advanced versus emerging innovation systems.

The remainder of the paper is organized as follows. Section II reviews the related literature. Section III introduces the Chinese job postings and patent data and describes how we measure technologies and diffusion. Section IV presents the empirical framework. Section V documents the results, focusing on baseline diffusion, heterogeneity by skills and technology characteristics, and the effects of external shocks. Section VI discusses policy implications, and Section VII concludes.

II. Related Literature

We connect to four strands of research: (i) technology diffusion and adoption, (ii) the geography of innovation and place-based policy, (iii) labor markets, skills, and postings data, and (iv) emerging-market innovation and international linkages.

Technology diffusion and adoption. Seminal studies document how technologies spread and raise productivity, with adoption shaped by local conditions and absorptive capacity (Griliches, 1957; Comin and Hobijn, 2004; Keller, 2004). Patent-based evidence highlights the role of spillovers and geographic localization in diffusion (Jaffe et al., 1993; Hall et al.,

2001). More recently, Kalyani et al. (2025) identify new technologies from U.S. patent text and trace their diffusion across regions and occupations, documenting that adoption begins in high-skill, high-wage locations and gradually broadens to lower-skill jobs and peripheral regions. These studies provide a baseline for diffusion in frontier economies. Our paper complements this work by studying diffusion in an emerging-market context, where shallow talent pools, dependence on foreign knowledge, and exposure to shocks generate distinctive frictions.

Geography of innovation and place-based policy. Innovation is spatially clustered; localized spillovers and agglomeration economies advantage dense regions and skill hubs (Jaffe et al., 1993; Combes et al., 2011; Moretti, 2012). Place-based policies can alter local trajectories by shifting infrastructure, human capital, and firm entry (Kline and Moretti, 2014). Our results speak directly to these themes but in an emerging-market setting. The speed of diffusion in China depends on workforce composition (low-skill versus high-skill technologies), technology position (upstream versus downstream; domestic versus U.S.-dependent), and shocks (COVID-19 mobility restrictions; U.S. sanctions). These findings suggest that effective place-based strategies in emerging economies must jointly address human-capital bottlenecks, upstream capabilities, and exposure to global constraints.

Labor markets, skills, and postings data. Job-vacancy text has become a central lens for studying technological change, capturing shifts in tasks and skills demanded by firms. Hershbein and Kahn (2018) show that recessions accelerate adoption of routine-biased technologies. Related work links automation and task content to labor-market outcomes (Acemoglu and Restrepo, 2022), and measures the spread of artificial intelligence specifically through online vacancies (Alonso et al., 2022). Deming (2017) highlight the rising importance of social skills in labor markets. We extend this approach to an emerging-market context, constructing a national postings dataset for China and using large language models to classify occupations and skills into standard taxonomies. By linking postings to patent-based

technologies, we trace diffusion across space and skill groups, showing that the availability of talent is a first-order determinant of diffusion speed.

Emerging-market innovation and international linkages. A growing literature studies innovation outside frontier economies. Ayyagari et al. (2011) analyze around 19,000 firms across 47 developing economies and show that access to external finance, governance structures, and exposure to foreign competition significantly affect firm innovation. Gorodnichenko et al. (2010) examine 27 emerging markets and demonstrate the role of globalization—via trade linkages, foreign competition, and supply chains—in driving domestic innovation. For the China case, Hu and Jefferson (2009), König et al. (2022), and Han et al. (2024) explore patent expansion, imitation-to-innovation transition, and U.S.–China technological interdependence. Our study builds on this literature by tracing how technologies diffuse across cities and occupations in China, and by emphasizing how labor-market frictions, technology position, and external dependence shape diffusion dynamics.

Taken together, these literatures highlight the importance of understanding how new technologies diffuse across space, skills, and institutional contexts. Our contribution is threefold. First, we provide new evidence from an emerging-market setting, where rapid innovation has occurred under a state-led model and in the shadow of global technological interdependence. Second, we construct a novel dataset that links patents with more than 31 million job postings, allowing us to measure diffusion not only across regions but also across occupations and skill groups. Third, we extend existing work by showing how emerging-market frictions—thin skill pools, dependence on foreign knowledge, and vulnerability to shocks—systematically shape the diffusion process. In this way, our analysis complements the U.S.-focused findings of Kalyani et al. (2025) and connects the literature on technology diffusion with research on emerging-market innovation and global linkages.

III. Data and Measurement

Our analysis draws on two newly assembled data sources: a large-scale dataset of job postings from major Chinese online platforms, and patent data from the China National Intellectual Property Administration (CNIPA). The former allows us to observe labor demand for specific skills and technologies, while the latter provides a systematic record of innovative output. By integrating the two, we can measure how technologies originate in patents and diffuse into the labor market across cities and occupations.

A. Chinese Labor Market Data

We build a new dataset of job postings from three major online recruiting platforms in China: 51job, Lagou, and Liepin. The data cover July 2016 to June 2021 and report job titles, job descriptions, employer names, and release dates.[1] Using multiple platforms is important because it captures both traditional white-collar jobs and newer digital-sector jobs.

A key challenge is duplication. Employers often repost the same vacancy on the same platform or across different platforms to attract more applicants. To handle this, we define two postings by the same employer as duplicates if they have the same job title and description and are posted within one month. We keep only the first posting in such cases. After removing duplicates, the dataset contains 30,970,360 unique postings. Finally, we standardize employer names by linking them to a reference list of firm identifiers to ensure consistency across postings.

Occupational classification. Classifying postings into the Chinese National Occupation Classification Code (NOCC, 2022) is challenging because job titles are often short, platform conventions differ, and many titles have multiple meanings across industries. To address

[1] We restrict the sample to this period because credible and consistent coverage of job postings is available for all three platforms beginning in mid-2016.

these challenges, we combine a large language model with retrieval-augmented generation (RAG).[2] We use Qwen, an open-source first-tier Chinese LLM, to extract the main responsibilities from each job description and assign the posting to the closest NOCC *minor* class.[3]

RAG improves the accuracy of this process in three ways. First, it ensures that the model uses the official and most up-to-date NOCC definitions, which may not be fully encoded in its training data. Second, by retrieving and displaying these definitions alongside each posting, it helps the model disambiguate firm-specific or non-standard job titles (including synonyms and abbreviations) and map them to standardized NOCC categories. Third, by narrowing the choice set to a small list of candidate classes and their official descriptions, RAG constrains the model's output space, thereby reducing off-taxonomy errors and improving consistency across platforms. In practice, for each posting we retrieve a shortlist of relevant NOCC minor classes and their task descriptions, and then prompt Qwen to select the best match based on the extracted responsibilities. Details, prompt templates, and examples are provided in the Appendix A.

Skill classification. Standardizing skills is equally challenging. In the raw postings, skill requirements are recorded in unstructured free text, often mixing technical skills, soft skills, and firm-specific terminology. To convert these into a structured taxonomy, we again combine Qwen with RAG. We use Qwen to extract skill-related phrases from each posting and then classify them into the Burning Glass taxonomy, which defines 31 major classes, 442 minor classes, and over 32,000 detailed skills. We focus on the *minor* class level as our main unit of analysis.[4]

RAG plays the same role here as in the occupation classification: it grounds the model in

[2]RAG (retrieval-augmented generation) is a technique that supplements an LLM's prompts with information retrieved from an external knowledge base. This allows the model to incorporate domain-specific or updated information that may not be fully represented in its training data (see Gao et al., 2023).

[3]We compared classifications at the middle, minor, and unit levels by random sampling and manual checks. The minor-class level performed best, being fine-grained enough to capture occupational detail while remaining robust to noisy titles and descriptions.

[4]As with occupations, we validated the choice of classification level by manually checking random samples across different levels. The minor-class level offered the best balance: detailed enough to differentiate between types of skills, yet robust to noise and idiosyncratic phrasing in postings.

the official taxonomy, helps disambiguate noisy or colloquial terms, and improves consistency across platforms. In practice, for each extracted raw skill we retrieve a shortlist of relevant Burning Glass minor classes and their descriptions, and then prompt Qwen to assign the best match. Details of the prompt design, validation procedure, and examples are provided in the Appendix A.

Labor-market patterns. Table I compares the frequency of major skill categories in the Chinese dataset to the U.S.-based Burning Glass postings. Communication, leadership, and personal attributes rank among the top skills in both datasets, underscoring the universal value of soft skills. However, important differences emerge. Chinese postings place greater emphasis on design (0.17 vs. 0.04), engineering (0.14 vs. 0.10), and education and training (0.17 vs. 0.11), consistent with the country's large manufacturing base and rapid industrial upgrading. In contrast, U.S. postings feature higher demand for health care (0.28 vs. 0.04), customer and client support (0.30 vs. 0.10), and law and compliance (0.31 vs. 0.10), reflecting structural differences in the two economies.

[Insert Table I here]

B. Patent Data and Technology Identification

We complement the job postings with comprehensive patent data from CNIPA, covering all patents granted between 2000 and 2022. Restricting to granted patents ensures that only inventions judged to be novel and non-obvious are included. This data allows us to identify new technologies at the point of invention.

Extracting candidate technologies. We process patent titles and abstracts using the Jieba Chinese segmentation tool to extract bigrams (two-word phrases). After removing stopwords and rare terms (appearing in fewer than 1,000 patents), we obtain 25,078 candidate technology bigrams. To filter out general or nonsensical phrases, we prompt GPT-4 to label

whether each bigram represents a technology-related term. Human reviewers validate and refine this output, yielding a final set of 7,836 distinct technologies.

Linking to labor demand. Patents identify what is being invented, while job postings reveal what is being adopted in the labor market. To focus on economically relevant technologies, we retain only those that appear in both patents and postings, and are mentioned in at least 100 postings. This restriction yields 3,269 technologies for analysis.

Examples. Table II lists the thirty most frequent technologies in Chinese postings. Software, operating systems, data analysis, control algorithms, and automation dominate, reflecting the centrality of digital technologies. At the same time, architecture, electronics, and database management also appear, underscoring the integration of IT with industrial applications.

[Insert Table II here.]

C. Technology Characteristics and External Factors

We construct several measures to describe the characteristics of each technology. One dimension is whether a technology is *foundational* (upstream) or *applied* (downstream). We use patent citations to make this distinction: a technology is classified as upstream if it frequently appears in patents that are cited by later patents, and downstream if it mainly appears in patents that cite others.

A second dimension is international dependence. Following Han et al. (2024), we measure China's reliance on U.S. knowledge by computing two citation ratios: the ratio of Chinese patents citing U.S. versus Chinese prior art, and the ratio of U.S. patents citing Chinese versus U.S. prior art. A larger gap between these ratios indicates greater Chinese dependence on U.S. prior art.

In addition to these technology characteristics, we also consider two external shocks. The first is the COVID-19 pandemic, coded as all quarters after 2020Q1, when mobility restrictions were widely introduced. The second is U.S. sanctions, measured using the Bureau of Industry and Security (BIS) Entity List. We identify Chinese firms that appear on this list and link them to the technologies in which they are major patent producers.

D. Discussion

The combined dataset highlights several distinctive features of China's innovation and labor market. Compared to the United States, Chinese postings are more concentrated and place greater emphasis on design, engineering, and training skills, reflecting the country's industrial structure. The most salient technologies are digital and automation-related but closely tied to applied use cases, underscoring the integration of IT capabilities with manufacturing and services.

More broadly, the dataset provides a new lens for studying technology diffusion. By linking large-scale job postings with patent data, we can observe how technologies move from invention to adoption across cities and occupations. This perspective is particularly valuable in China, where rapid innovation has occurred under a state-led model and in the context of deep global interdependence. The data thus allow us to document not only broad diffusion patterns but also how they vary with workforce composition, technology characteristics, and external shocks.

IV. Empirical Framework

Our goal is to measure how new technologies diffuse across Chinese cities and occupations. We capture diffusion by examining how the spatial concentration of technology-specific job postings changes over time.

A. *Measuring Diffusion*

Following Kalyani et al. (2025), we adopt comparable measures of spatial concentration to benchmark our results against the United States, while applying them to China as an emerging-market case. For technology τ, city c, and quarter t, define

$$\text{NormShare}_{c,\tau,t} = \frac{\dfrac{\text{Postings}^{\text{tech}}_{c,\tau,t}}{\sum_{c'} \text{Postings}^{\text{tech}}_{c',\tau,t}}}{\dfrac{\text{Postings}^{\text{all}}_{c,t}}{\sum_{c'} \text{Postings}^{\text{all}}_{c',t}}}.$$

If postings mentioning a technology are exactly proportional to overall job activity, then $\text{NormShare}_{c,\tau,t} = 1$ in every city. Values above (below) one indicate that the technology is over-represented (under-represented) relative to a city's share of all labor demand.[5]

Figure 1 plots the geographic distribution of technology exposure across Chinese cities. Darker shading indicates higher exposure. Adoption is concentrated in a few hubs such as Beijing, Shanghai, and Shenzhen, while many interior regions have very low exposure. This pattern is similar to the United States, where Kalyani et al. (2025) document that the most economically significant new technologies originate in a handful of pioneer regions such as Silicon Valley and the Northeast Corridor.

[Insert Figure 1 here.]

We then construct two measures of concentration based on normalized shares. The first is the coefficient of variation (COV) across cities:

$$COV_{t,\tau} = \frac{\text{sd}\big(\text{NormShare}_{c,\tau,t}\big)}{\text{mean}\big(\text{NormShare}_{c,\tau,t}\big)}.$$

A high value of $COV_{t,\tau}$ indicates that postings are concentrated in a few locations, while a decline in $COV_{t,\tau}$ suggests that adoption is spreading more evenly across space.

[5] As emphasized by Kalyani et al. (2025), this normalization adjusts for city size and platform coverage, making technologies comparable across regions and time.

As an alternative measure, we compute the mean normalized share of postings in the top five cities, denoted $Top5_{t,\tau}$. A fall in $Top5_{t,\tau}$ similarly signals that diffusion is broadening beyond the initial centers of activity. Both measures are scale-free and comparable across technologies.

B. Baseline Specification

To quantify broadening over time, we estimate the following regression at the technology τ by quarter t level:

$$COV_{t,\tau} = \alpha_0 + \beta_{RB}(t - t_{0,\tau}) + \delta_\tau + \epsilon_{t,\tau}, \tag{1}$$

where $(t - t_{0,\tau})$ is the number of quarters since technology τ first appears in the postings data, and δ_τ are technology fixed effects.

The coefficient β_{RB} measures the *rate of regional broadening (RB)*. A negative β_{RB} indicates that concentration declines over time, meaning the technology diffuses more broadly across cities. In the U.S., Kalyani et al. (2025) estimate very slow broadening. In our setting, the same framework allows us to test whether emerging-market frictions accelerate or slow the pace of diffusion.

C. Extensions

The baseline captures the average diffusion rate across technologies. To examine heterogeneity, we augment equation (1) with interactions corresponding to the three emerging-market frictions highlighted in the introduction: shallow skill pools, the distinction between foundational and applied technologies, and reliance on foreign knowledge, as well as exposure to external shocks.

Human-capital constraints. We first compare diffusion across high-skill and low-skill technologies, based on occupational classifications. This allows us to test whether application-oriented technologies linked to widely available, lower-skill labor diffuse more rapidly than

those requiring specialized human capital. If skill bottlenecks constrain adoption, we expect diffusion to be slower for high-skill jobs.

Foundational versus applied technologies. Next, we distinguish between foundational (upstream) and applied (downstream) technologies using patent citation patterns. Upstream technologies typically require deeper knowledge bases and complementary capabilities, while downstream applications are easier to adopt. The framework therefore allows us to test whether technologies that are more upstream diffuse more slowly across Chinese cities.

Dependence on foreign knowledge. We then examine whether reliance on U.S. prior art hampers diffusion. Following Han et al. (2024), we use patent citations to measure international dependence and interact this with time since a technology's entry. If foreign dependence is a binding friction in an emerging market, diffusion should be slower for technologies rooted in U.S. knowledge.

External shocks. Finally, we study whether diffusion patterns are altered by major shocks. We focus on two events. The first is the COVID-19 pandemic, coded as an indicator equal to one for quarters after 2020Q1, when widespread mobility restrictions were introduced. The second is U.S. sanctions, measured using the Bureau of Industry and Security Entity List, which identifies Chinese entities subject to export restrictions. These shocks provide a way to examine how unexpected disruptions to mobility or foreign access reshape diffusion trajectories.

Details of variable construction and the specific regression equations are provided in Section V when we present the results.

V. Results

A. *Baseline Diffusion*

We begin with the baseline regression in equation (1). Table III reports estimates using both concentration measures: the coefficient of variation (COV) of normalized shares (columns 1–2) and the share of postings in the top five cities ($Top5_{t,\tau}$, columns 3–4). Odd-numbered columns exclude technology fixed effects, while even-numbered columns include them.

[Insert Table III here.]

Across all specifications, the coefficient on regional broadening, β_{RB}, is negative and highly significant. In column (2), which includes technology fixed effects, the estimate is $\hat{\beta}_{RB} = -0.030$ (s.e. 0.002). This implies that the COV falls by 0.030 units per quarter after a technology first appears, consistent with steady geographic diffusion. Columns (3) and (4) show similar results for the $Top5_{t,\tau}$ measure: the share of postings in the top five cities declines by about 0.005 per quarter.

These results indicate that new technologies in China are not confined to a few hubs but diffuse across regions over time. The pattern is similar to the findings of Kalyani et al. (2025) for the United States, though in a very different institutional context.

B. *Human-Capital Constraints*

We next examine how diffusion varies with skill intensity, reflecting constraints in the availability of specialized talent. We classify postings into high-, medium-, and low-skill groups based on their dependence on knowledge-intensive tasks. For example, IT and Software Service Professionals are classified as high-skill occupations, while Sales or Administrative Staff are classified as low-skill occupations. This classification was generated by GPT-4 and refined by human review.

We implement this analysis in two ways.

(i) Split-sample specification. For $s \in \{H, L\}$ (high- and low-skill), we estimate

$$COV^{s}_{t,\tau} \;=\; \alpha^{s}_{0} \;+\; \beta^{s}_{RB}(t - t_{0,\tau}) \;+\; \delta_{\tau} \;+\; \varepsilon^{s}_{t,\tau}. \tag{2}$$

This allows the rate of diffusion, β^{s}_{RB}, to differ across high- and low-skill technologies.

(ii) Interaction specification. Let $I^{\tau,t}_{\text{low}} = 1$ when postings for technology τ in quarter t are classified as low-skill (0 otherwise). We then estimate

$$COV_{t,\tau} \;=\; \alpha_{0} \;+\; \beta_{RB}(t - t_{0,\tau}) \;+\; \beta_{\text{low}}\, I^{\tau,t}_{\text{low}} \;+\; \beta_{RB,\text{low}}\, I^{\tau,t}_{\text{low}} \cdot (t - t_{0,\tau}) \;+\; \delta_{\tau} \;+\; \varepsilon_{t,\tau}. \tag{3}$$

Here, $\beta_{RB,\text{low}}$ captures the difference in diffusion speed between low- and high-skill technologies.

Table IV reports the results. In column (2), the slope for low-skill postings is $\hat{\beta}^{L}_{RB} = -0.036$ (s.e. 0.002), compared to $\hat{\beta}^{H}_{RB} = -0.022$ (s.e. 0.002) for high-skill postings. The interaction specification confirms this pattern: $\hat{\beta}_{RB,\text{low}} = -0.015$ (s.e. 0.001), implying faster diffusion of low-skill technologies.

[Insert Table IV here.]

These results indicate that technologies linked to low-skill, application-oriented jobs diffuse across cities more quickly than those requiring high-skill labor. This pattern reflects the broader availability of low-skill workers across regions, while high-skill technologies depend on concentrated R&D facilities and specialized talent.

C. Foundational versus Applied Technologies

We then explore how diffusion varies with the position of a technology in the innovation chain. Foundational (upstream) technologies typically require deeper knowledge bases, while applied (downstream) technologies are closer to end-user applications.

Formally, let P_τ denote the set of patents that contain the technology bigram τ. Define P_τ^U as the set of patents *cited by* P_τ (backward citations), and P_τ^D as the set of patents that *cite* P_τ (forward citations). We construct

$$UpDownRatio_\tau \;=\; \frac{|P_\tau^D|+1}{|P_\tau^U|+1},$$

so that larger values indicate a technology that is more *foundational/upstream*. We also define a binary indicator

$$IsUpstream_\tau \;=\; \mathbb{1}\{UpDownRatio_\tau > 1\}.$$

Table V lists examples of the most foundational and most applied technologies in our sample. Upstream technologies such as *Fourier Transform*, *Compound Semiconductor*, and *Programmable Logic Controller* provide core principles that underpin other advancements, while downstream technologies such as *Artificial Intelligence Technology*, *New Energy Vehicles*, and *Wearable Devices* are application-oriented and closer to end users.

[Insert Table V here.]

To test whether foundational and applied technologies diffuse at different speeds, we estimate:

$$COV_{t,\tau} \;=\; \alpha_0 \;+\; \beta_{RB}(t-t_{0,\tau}) \;+\; \beta_{Up,RB}\,Upstream_\tau \cdot (t-t_{0,\tau}) \;+\; \delta_\tau \;+\; \varepsilon_{t,\tau}, \qquad (4)$$

and analogously with $Top5_{t,\tau}$ as the dependent variable. Here, $Upstream_\tau$ takes either $UpDownRatio_\tau$ or $IsUpstream_\tau$, so that larger values represent more foundational technologies. A positive $\beta_{Up,RB}$ indicates slower diffusion for upstream technologies.

[Insert Table VI here.]

Table VI shows that $\hat{\beta}_{Up,RB}$ is positive and significant across specifications, implying that foundational technologies diffuse more slowly. This is consistent with their complexity and reliance on specialized expertise, which make them more likely to remain concentrated in major hubs.

D. Dependence on Foreign Knowledge

We next examine whether diffusion is slowed when technologies depend more heavily on foreign (particularly U.S.) knowledge. Greater dependence may hinder adoption if diffusion is limited by access to foreign intellectual property, complementary inputs, or absorptive capacity. Following Han et al. (2024), we use patent citations to construct measures of international dependence.

Let $n_{c,u,t}$ denote the number of citations that Chinese patents make to U.S. patents at time t, and $n_{c,c,t}$ the citations that Chinese patents make to other Chinese patents. Similarly, let $n_{u,c,t}$ be the citations U.S. patents make to Chinese patents, and $n_{u,u,t}$ the citations U.S. patents make to other U.S. patents. Let $x_{c,t}$ and $x_{u,t}$ be the cumulative number of Chinese and U.S. patents granted up to time t. We then define

$$p_{c,u,t} = \frac{n_{c,u,t}/x_{u,t}}{n_{c,c,t}/x_{c,t}}, \qquad p_{u,c,t} = \frac{n_{u,c,t}/x_{c,t}}{n_{u,u,t}/x_{u,t}}.$$

Here, $p_{c,u,t}$ is the relative propensity of Chinese patents to cite U.S. prior art, and $p_{u,c,t}$ is the relative propensity of U.S. patents to cite Chinese prior art.

We define a binary indicator

$$Dep_{t,\tau} = \mathbb{1}\{\, p_{c,u,t} - p_{u,c,t} > 0 \,\},$$

which equals one if Chinese patents rely more heavily on U.S. prior art than vice versa, indicating dependence on U.S. technology.

Table VII lists examples of technologies with high and low dependence. Highly depen-

dent technologies include semiconductors, electronic circuits, and telecommunication systems, while less dependent ones include new energy vehicles, wind turbines, and breeding technologies.

[Insert Table Table VII here.]

To test whether dependence slows diffusion, we estimate

$$COV_{t,\tau} \;=\; \alpha_0 \;+\; \beta_{RB}(t - t_{0,\tau}) \;+\; \beta_{Dep}\, Dep_{t,\tau} \;+\; \beta_{Dep,RB}\, Dep_{t,\tau} \cdot (t - t_{0,\tau}) \;+\; \delta_\tau \;+\; \varepsilon_{t,\tau}, \quad (5)$$

and analogously with $Top5_{t,\tau}$ as the dependent variable. A positive $\beta_{Dep,RB}$ indicates that U.S.-dependent technologies diffuse more slowly across Chinese cities.

Table VIII shows that dependence on U.S. prior art is associated with significantly slower regional broadening, consistent with barriers to adoption when technologies are rooted in foreign knowledge bases.

[Insert Table Table VIII here.]

E. External Shocks

We next examine how diffusion responded to two major shocks: the COVID-19 pandemic and U.S. sanctions imposed through the Bureau of Industry and Security (BIS) Entity List.

COVID-19 pandemic. The pandemic disrupted mobility and economic activity, with implications for technologies tied to in-person work. Estimates in Table IX show that diffusion slowed after 2020Q1: technologies became more concentrated and the rate of regional broadening declined. This effect was especially pronounced for low-skill, application-oriented technologies, which rely on physical presence. High-skill technologies, more knowledge-intensive and often adaptable to remote collaboration, were less affected. The full regression specification and variable construction are provided in Appendix B.A.

[Insert Table IX here.]

U.S. sanctions. Sanctions restricted access to U.S. technologies and inputs but also spurred domestic substitution. Table X shows that diffusion accelerated once a technology became linked to a sanctioned entity: the estimated slope of broadening became more negative. This suggests that sanctions acted as a catalyst for localization, pushing firms and regions to adopt related technologies more broadly within China. Details of the estimation framework and robustness checks are reported in Appendix B.B.

[Insert Table X here.]

F. Summary

Taken together, our results highlight how diffusion in China is shaped by emerging-market frictions. Technologies begin concentrated in a few hubs but broaden over time, as in the United States. Yet the rate of diffusion depends systematically on three factors: the availability of specialized human capital, whether technologies are foundational or applied, and reliance on foreign knowledge. External shocks further shift diffusion: COVID slowed it, especially for low-skill technologies, while U.S. sanctions accelerated it through substitution and localization. These findings show how the Chinese case complements U.S. evidence by illustrating how emerging-market frictions condition the diffusion of new technologies.

VI. Policy Discussion

Our findings have several implications for place-based policies, labor-market development, and strategies to manage global innovation in emerging economies.

A. Human-capital constraints and regional policy

The evidence that low-skill technologies diffuse more rapidly than high-skill ones reflects the uneven distribution of talent across Chinese cities. Diffusion of complex, research-intensive technologies depends on local human-capital endowments and R&D infrastructure,

which remain concentrated in major hubs. This underscores the importance of place-based policies that expand advanced training, graduate education, and research capacity outside established centers. By broadening the geographic distribution of high-skill labor, policymakers can accelerate the spread of foundational technologies and reduce persistent regional disparities.

B. Foundational versus applied technologies and absorptive capacity

Foundational technologies diffuse more slowly than applied ones, consistent with their complexity and greater reliance on specialized expertise. This suggests that investments in absorptive capacity—such as technical universities, research institutes, and advanced manufacturing facilities—are critical for ensuring that peripheral regions can adopt upstream technologies rather than being confined to downstream applications. Without such investments, emerging economies risk locking regions into a limited segment of the innovation chain.

C. Foreign dependence, shocks, and resilience

Dependence on U.S. prior art slows diffusion, while sanctions unexpectedly accelerated it by spurring substitution and localization. These findings highlight the dual nature of global interdependence: foreign knowledge is a vital input for innovation, but reliance on it can also create vulnerabilities. For policymakers, this points to a trade-off between openness and resilience. Reducing excessive dependence on foreign knowledge may safeguard against geopolitical shocks, while international collaboration remains essential for accessing the frontier.

D. Broader implications

Taken together, the results show that technology diffusion in emerging markets is shaped by the interaction of human capital, technology complexity, and international linkages. Place-based policies that focus only on infrastructure or subsidies are unlikely to succeed without parallel investments in skills and absorptive capacity. Similarly, global strategies must balance the benefits of integration with the vulnerabilities that dependence creates. The Chinese case illustrates how emerging-market frictions shape diffusion, offering lessons for other economies seeking to manage the geography of innovation under similar constraints.

VII. Conclusion

This paper studies how new technologies diffuse in an emerging-market setting, using China as a case. We construct a new dataset linking more than 30 million online job postings to nearly two decades of patenting activity, allowing us to trace how 3,269 patented technologies move from invention to adoption in the labor market. This framework provides a novel view of the geography of diffusion in an economy that has become a central player in global innovation.

Three main findings emerge. First, consistent with evidence for the United States (Kalyani et al., 2025), new technologies in China begin highly concentrated in a few hubs but gradually broaden across regions. Second, the pace of diffusion varies systematically with emerging-market frictions. Technologies tied to widely available, lower-skill labor and downstream applications diffuse relatively quickly, while those requiring scarce high-skill talent, foundational knowledge, or heavy reliance on U.S. prior art diffuse more slowly. Third, external shocks matter. The COVID-19 pandemic slowed diffusion, particularly for occupations dependent on physical presence, whereas U.S. sanctions unexpectedly accelerated diffusion by catalyzing substitution and localization.

Overall, our findings show that the diffusion of new technologies is shaped not only by

market incentives but also by the interaction of three structural frictions—human-capital constraints, technological complexity, and foreign dependence—together with exposure to shocks. The Chinese case complements U.S. evidence by illustrating how these frictions operate in an emerging innovation hub where state policy and global interdependence play central roles. More broadly, the results underscore the importance of workforce development, absorptive capacity, and resilience for managing the geography of innovation in emerging economies.

REFERENCES

Acemoglu, Daron, and Pascual Restrepo, 2022, Tasks, automation, and the rise in us wage inequality, *Econometrica* 90, 1973–2016.

Alonso, Crespo, Lee Branstetter, et al., 2022, Tracking the diffusion of artificial intelligence with online job postings, *Labour Economics* 76, 102157.

Ayyagari, Meghana, Asli Demirgüç-Kunt, and Vojislav Maksimovic, 2011, Firm innovation in emerging markets: The role of finance, governance, and competition, *Journal of Financial and Quantitative Analysis* 46, 1545–1580.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, 2011, The identification of agglomeration economies, *Journal of Economic Geography* 11, 253–266.

Comin, Diego, and Bart Hobijn, 2004, Cross-country technology adoption: Making the theories face the facts, *Journal of Monetary Economics* 51, 39–83.

Deming, David J., 2017, The growing importance of social skills in the labor market, *Quarterly Journal of Economics* 132, 1593–1640.

Gao, Yunfan, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang, 2023, Retrieval-augmented generation for large language models: A survey, Preprint, *arXiv*:2312.10997.

Gorodnichenko, Yuriy, Jan Svejnar, and Katherine Terrell, 2010, Globalization and innovation in emerging markets, *American Economic Journal: Macroeconomics* 2, 194–226.

Griliches, Zvi, 1957, Hybrid corn: An exploration in the economics of technological change, *Econometrica* 25, 501–522.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg, 2001, The nber patent citation data file: Lessons, insights and methodological tools, Technical Report 8498, National Bureau of Economic Research.

Han, Pengfei, Wei Jiang, and Danqing Mei, 2024, Mapping u.s.–china technology decoupling: Policies, innovation, and firm performance, *Management Science* 70, 8386–8413.

Hershbein, Brad, and Lisa B. Kahn, 2018, Do recessions accelerate routine-biased technological change? evidence from vacancy postings, *American Economic Review* 108, 1737–1772.

Hu, Albert Guangzhou, and Gary H. Jefferson, 2009, A great wall of patents: What is behind china's recent patent explosion?, *Journal of Development Economics* 90, 57–68.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson, 1993, Geographic localization of knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics* 108, 577–598.

Kalyani, Aakash, Nicholas Bloom, Marcela Carvalho, Tarek A. Hassan, Josh Lerner, and Ahmed Tahoun, 2025, The diffusion of new technologies, *Quarterly Journal of Economics* 140, 1299–1365.

Keller, Wolfgang, 2004, International technology diffusion, *Journal of Economic Literature* 42, 752–782.

Kline, Patrick, and Enrico Moretti, 2014, Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority, *Quarterly Journal of Economics* 129, 275–331.

König, Michael, Zheng Song, Kjetil Storesletten, and Fabrizio Zilibotti, 2022, From imitation to innovation: Where is all that chinese r&d going?, *Econometrica* 90, 1615–1654.

Moretti, Enrico, 2012, *The New Geography of Jobs* (Houghton Mifflin Harcourt).

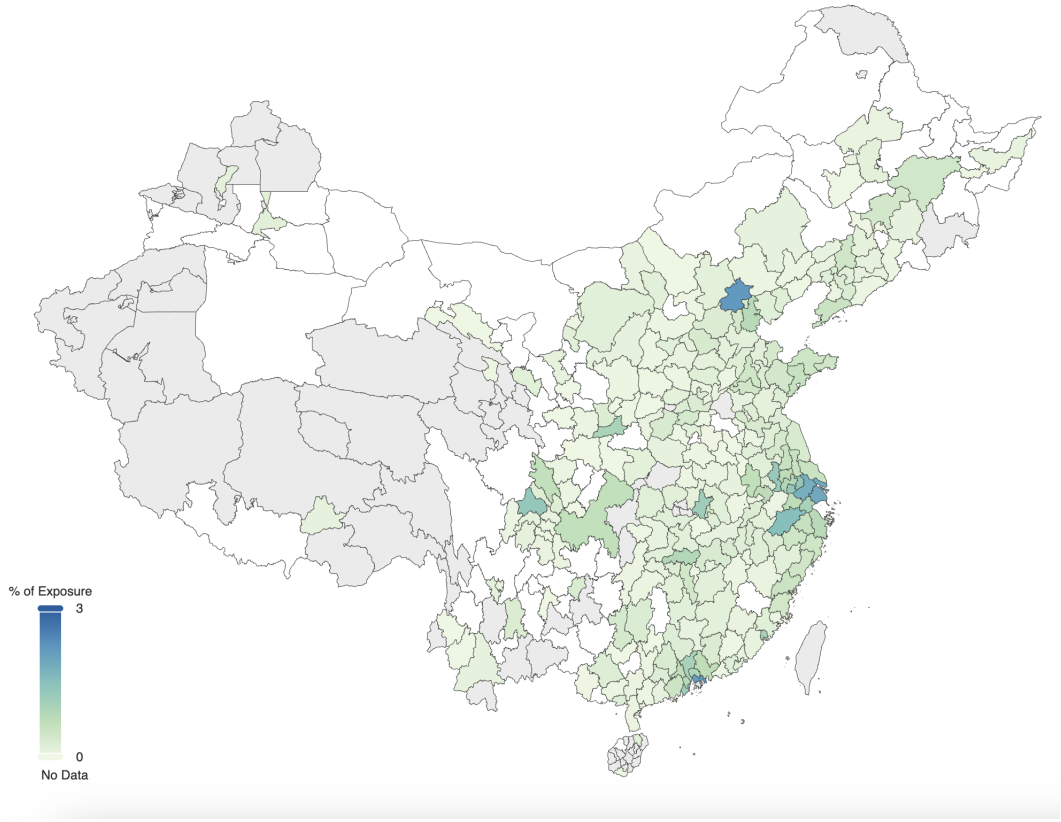


Figure 1. Geographic distribution of technology exposure across Chinese cities. Darker shading indicates a higher share of exposure. Exposure is concentrated in a few hubs (e.g., Beijing, Shanghai, Shenzhen) with much lower levels elsewhere.

Table I

Major Skills Frequency in Chinese Job Postings and Burning Glass

The table compares the frequency of major skill categories in Chinese job postings (2016–2021) with the U.S. Burning Glass dataset. Frequencies are shares of all postings within each dataset.

Skill Name	Chinese Dataset Frequency	Burning Glass Frequency
Media and Communications	0.61	0.51
Physical and Inherent Abilities	0.58	0.61
Business	0.35	0.50
Information Technology	0.35	0.34
Marketing and Public Relations	0.24	0.12
Sales	0.19	0.26
Education and Training	0.17	0.11
Design	0.17	0.04
Administration	0.16	0.27
Engineering	0.14	0.10
Finance	0.13	0.24
Analysis	0.11	0.16
Customer and Client Support	0.10	0.30
Law, Regulation, and Compliance	0.10	0.31
Manufacturing and Production	0.08	0.11
Hospitality and Food Services	0.07	0.09
Transportation, Supply Chain, and Logistics	0.06	0.20
Architecture and Construction	0.06	0.09
Human Resources	0.06	0.07
Maintenance, Repair, and Facility Services	0.05	0.15
Economics, Policy, and Social Studies	0.05	0.04
Public Safety and National Security	0.04	0.07
Health Care	0.04	0.28
Property and Real Estate	0.03	0.02
Science and Research	0.03	0.12
Social and Human Services	0.02	0.02
Personal Care and Services	0.02	0.01
Energy and Utilities	0.02	0.02
Environment	0.02	0.02
Performing Arts, Sports, and Recreation	0.01	0.02
Agriculture, Horticulture, and Landscaping	0.01	0.01

Table II

Top 30 New Technologies in China

The table lists the 30 most common new technologies identified from patents and job postings in China. Technologies are grouped into three illustrative columns for display purposes.

Software	Network	Motivation
System	Internet	Operating System
Data	Programming	Model
Computer	Module	Video
Test	Program	Communication
Control	Algorithm	Data Structure
Equipment	Automation	Channel
Database	Craftsmanship	Calculation
Architecture	Code	Logic
Data Analysis	Electronics	Computer

Table III

Technology Diffusion Across Cities: Baseline

This table reports estimates of equation (1) at the technology-by-quarter level. The dependent variables are the coefficient of variation (COV) of normalized shares across cities and the mean normalized share in the top five cities ($Top5_{t,\tau}$). Columns (1) and (3) exclude technology fixed effects; columns (2) and (4) include them. Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	COV of Normalized Shares		Top-5 City Share	
	(1)	(2)	(3)	(4)
β_{RB} (rate of regional broadening)	-0.007**	-0.030***	-0.004***	-0.005***
	(0.003)	(0.002)	(0.000)	(0.000)
α_0	9.229***	9.476***	0.829***	0.844***
	(0.036)	(0.024)	(0.002)	(0.001)
Technology FE	No	Yes	No	Yes
R^2	0.000	0.591	0.008	0.761
N	62,114	62,114	62,114	62,114

Table IV

Diffusion by Skill Intensity

This table reports estimates of equations (2) and (3) at the technology-by-quarter level. Columns (1)–(2) show split-sample estimates for high- and low-skill postings. Columns (3) and (6) add interactions with a low-skill indicator. Dependent variables are the coefficient of variation (COV) of normalized shares across cities and the mean normalized share in the top five cities ($Top5_{t,\tau}$). All models include technology fixed effects. Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	COV of Normalized Shares			Top-5 City Share		
	High-Skill	Low-Skill	All	High-Skill	Low-Skill	All
β_{RB}	-0.022***	-0.036***	-0.022***	-0.005***	-0.006***	-0.005***
	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)
β_{low}			-0.014**			-0.003***
			(0.006)			(0.000)
$\beta_{RB,\text{low}}$			-0.015***			-0.001***
			(0.001)			(0.000)
α_0	9.409***	9.396***	9.409***	0.846***	0.844***	0.846***
	(0.024)	(0.024)	(0.024)	(0.001)	(0.001)	(0.001)
Technology FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.599	0.591	0.606	0.762	0.758	0.766
N	62,114	62,114	124,228	62,114	62,114	124,228

Table V

Top 30 Upstream and Downstream Technologies

The table lists the thirty most upstream and thirty most downstream technologies identified from patent citation patterns. Upstream technologies are those that are cited more frequently by later patents, while downstream technologies are those that cite more prior patents and are closer to applications.

Top 30 Upstream Technologies		
Image	PC machine	Control the computer
Microcomputer	Monitoring Center	Search Engine
Telecommunication system	Compound Semiconductor	Air conditioner
Mobile phone	Wireless sensor	Programmable Logic Controller
Internet	Sensor network	Digital Signal Processor
Architecture	Liquid Crystal Display	Monitor
Camera	Microwave oven	Program Control
Client	Route selection	Digital Camera
Transceiver	Digital Data	Embedded system
IP network	Fourier Transform	Aerospace
Top 30 Downstream Technologies		
Storage medium	Display module	Equipment Technology
Computer equipment	Driver Component	Device Technology
Artificial Intelligence Technology	Networked device	Terminal equipment
Data processing technology	Air conditioning technology	Target audience
Precision Control	Additive Manufacturing	Computing device
Artificial Intelligence field	Data processing method	Model Training Method
Electronic equipment	Wearable Devices	Driving mechanism
Primary Information	Artificial Intelligence	Data processing field
New energy vehicles	Computer Technology Field	Wastewater treatment technology
Precise positioning	Autonomous driving	Auto parts

Table VI

Diffusion and Technology Position: Upstream vs. Downstream

This table reports estimates of equation (4) at the technology-by-quarter level. The dependent variables are the coefficient of variation (COV) of normalized shares across cities and the mean normalized share in the top five cities ($Top5_{t,\tau}$). $Upstream_\tau$ is measured either as a binary indicator $IsUpstream_\tau = \mathbb{1}\{UpDownRatio_\tau > 1\}$ or as the continuous ratio $UpDownRatio_\tau = (|P_\tau^D| + 1)/(|P_\tau^U| + 1)$. Larger values indicate more upstream (foundational) technologies. Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	COV of Normalized Shares		Top-5 City Share	
$Upstream_\tau$	$IsUpstream_\tau$	$UpDownRatio_\tau$	$IsUpstream_\tau$	$UpDownRatio_\tau$
$\beta_{Up,RB}$	0.022***	0.066***	0.002***	0.004***
	(0.005)	(0.013)	(0.000)	(0.001)
β_{RB}	-0.046***	-0.036***	-0.006***	-0.006***
	(0.004)	(0.002)	(0.000)	(0.000)
α_0	9.476***	9.476***	0.844***	0.844***
	(0.024)	(0.024)	(0.001)	(0.001)
Technology FE	Yes	Yes	Yes	Yes
R^2	0.591	0.591	0.761	0.761
N	62,114	62,114	62,114	62,114

Table VII

Technologies with High and Low U.S. Dependence

The table lists examples of technologies with high and low dependence on U.S. prior art, based on relative patent citation patterns. High-dependence technologies include semiconductors, circuits, and telecommunication systems. Low-dependence technologies are more rooted in domestic knowledge, such as new energy vehicles and renewable energy systems.

High U.S. Dependence Technologies	**Low U.S. Dependence Technologies**
Semiconductor	Biological Engineering Technology
Electronic circuit	New energy vehicles
Telecommunication system	Electric energy quality
Storage device	Information interaction
Etching process	Wind turbine generator set
Internal combustion engine	Model Training Method
Internet	3D Print Cloud
Imaging equipment	Breeding technology
Ink-jet printer	Anti-interference ability
Wireless device	Mechanical Processing Technology

Table VIII

Influence of U.S. Dependence on the Diffusion of New Technologies

This table reports estimates of equation (5) at the technology-by-quarter level. The dependent variables are the coefficient of variation (COV) of normalized shares across cities and the mean normalized share in the top five cities ($Top5_{t,\tau}$). $Dep_{t,\tau}$ equals one if Chinese patents for technology τ rely more heavily on U.S. prior art than vice versa. Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	COV of Normalized Shares	Top-5 City Share
β_{RB}	-0.042***	-0.006***
	(0.003)	(0.000)
β_{Dep}	-0.402***	-0.025***
	(0.058)	(0.003)
$\beta_{Dep,RB}$	0.028***	0.002***
	(0.004)	(0.000)
α_0	9.610***	0.853***
	(0.035)	(0.002)
Technology FE	Yes	Yes
R^2	0.595	0.763
N	59,299	59,299

Table IX

Influence of COVID-19 on the Diffusion of New Technologies

This table reports estimates of equations (B1) and (B2) at the technology-by-quarter level. The dependent variables are the coefficient of variation (COV) of normalized shares across cities and the mean normalized share in the top five cities ($Top5_{t,\tau}$). Columns (1) and (3) show baseline estimates without skill interactions, while columns (2) and (4) add interactions with a low-skill indicator. All models include technology fixed effects. Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	COV of Normalized Shares		Top-5 City Share	
	Baseline	With Skill Interactions	Baseline	With Skill Interactions
β_{COVID}	-3.608***	-2.603***	-0.278***	-0.220***
	(0.194)	(0.195)	(0.008)	(0.008)
β_{RB}	0.005	0.007**	-0.006***	-0.005***
	(0.004)	(0.003)	(0.000)	(0.000)
$\beta_{COVID,RB}$	0.171***	0.121***	0.016***	0.013***
	(0.012)	(0.012)	(0.000)	(0.000)
β_{low}		-0.050***		-0.003***
		(0.007)		(0.000)
$\beta_{RB,\text{low}}$		-0.007***		-0.001***
		(0.001)		(0.000)
$\beta_{COVID,\text{low}}$		-1.469***		-0.081***
		(0.053)		(0.002)
$\beta_{COVID,RB,\text{low}}$		0.075***		0.004***
		(0.003)		(0.000)
α_0	9.291***	9.256***	0.851***	0.852***
	(0.030)	(0.029)	(0.002)	(0.002)
Technology FE	Yes	Yes	Yes	Yes
R^2	0.595	0.609	0.765	0.769
N	62,114	124,228	62,114	124,228

Table X

Influence of U.S. Entity List Sanctions on the Diffusion of New Technologies

This table reports estimates of equation (B3) at the technology-by-quarter level. The dependent variables are the coefficient of variation (COV) of normalized shares across cities and the mean normalized share in the top five cities ($Top5_{t,\tau}$). $I_{t,\tau}^{Entity} = 1$ if a top patent-producing entity for technology τ is on the BIS Entity List in quarter t. All models include technology fixed effects. Standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	COV of Normalized Shares	Top-5 City Share
β_{RB}	-0.019***	-0.005***
	(0.005)	(0.000)
β_{Entity}	-0.040	0.004
	(0.101)	(0.005)
$\beta_{Entity,RB}$	-0.012**	-0.001***
	(0.006)	(0.000)
α_0	9.498***	0.841***
	(0.077)	(0.004)
Technology FE	Yes	Yes
R^2	0.591	0.761
N	62,114	62,114

Appendix A. Text Processing Methodology

This appendix describes the Retrieval-Augmented Generation (RAG) methodology used for classifying job titles and skills from raw job posting text. As discussed in Section III, a key challenge in our data processing pipeline is to accurately and consistently map unstructured, often noisy, information from job postings to standardized taxonomies. To address this, we developed a two-stage process that leverages a state-of-the-art open-source Large Language Model (LLM) at the time of research, the Qwen-2.0 model, augmented with external knowledge bases. This RAG approach enhances classification accuracy by grounding the model's responses in official, domain-specific information.

This process is generalizable and can be adapted for various text-structuring tasks. In our research, it is applied to two main applications: (1) Occupational Classification and (2) Skill Classification. Both applications share the same core RAG workflow, as illustrated below.

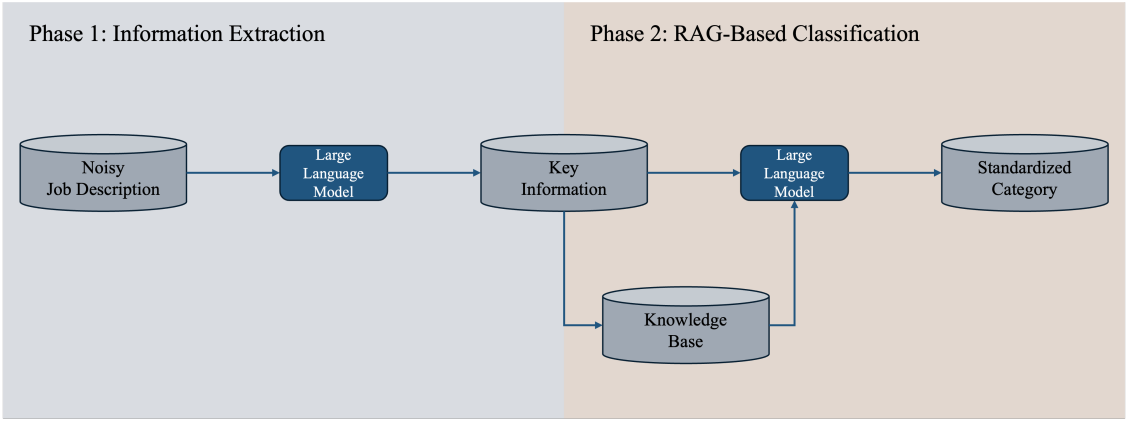


Figure A2. The RAG-Based Text Classification Workflow

General RAG Workflow

Our RAG-based classification workflow consists of the following four steps:

1. **Information Extraction:** For each raw job posting, the LLM performs an initial processing step to extract key information. For occupational classification, this involves

identifying the primary job responsibilities and tasks. For skill classification, this involves extracting all phrases that represent a specific skill. This step effectively cleans and structures the noisy input text.

2. **Candidate Retrieval:** The extracted key information (e.g., job responsibilities or a skill phrase) is used as a query to retrieve a set of relevant candidate entries from an external knowledge base. For occupations, the knowledge base is the Chinese National Occupation Classification Code (NOCC, 2022) handbook. For skills, it is the Burning Glass taxonomy. This step narrows the vast number of possible classifications to a small, highly relevant subset.

3. **Prompt Augmentation:** A detailed prompt is constructed for the LLM. This prompt includes the original job description, the extracted key information from Step 1, and the list of retrieved candidate classifications along with their official definitions from Step 2. This augmentation provides the model with critical context and constrains it to choose from a valid set of options.

4. **Final Classification:** The augmented prompt is fed to the LLM, which generates the final classification. By having official definitions and a candidate list in its context window, the model can accurately disambiguate non-standard job titles or firm-specific jargon and map them to the correct standardized code.

This workflow ensures that classifications are not only based on the LLM's general knowledge but are also aligned with the specific requirements of our target taxonomies, significantly reducing errors and improving consistency.

Application 1: Occupational Classification

The primary objective of this application is to map each job posting to the most appropriate minor class within the Chinese National Occupation Classification Code (NOCC,

2022). To achieve this, the process utilizes the official NOCC 2022 taxonomy, which serves as the external knowledge base and provides detailed descriptions for each occupational class.

Prompt Templates

All original inputs, prompts, and outputs have been translated from Chinese into English for ease of reading.[6] The following simplified templates illustrate the prompts used in the task. Placeholders are filled programmatically for each job posting.

Prompt for Key Responsibility Extraction.

```
You are an expert in human resources and occupational classification.
Your task is to accurately assign a job posting to the most suitable
NOCC 2022 minor class code. Please follow the instructions below.

Please extract the key responsibilities from the job posting
based on the job description:

{raw_job_description_text}

The extracted responsibilities are:
```

Prompt for Occupation Classification.

[6]The translation preserves the structure of the prompts but not the exact wording of the original Chinese inputs.

```
You are an expert in human resources and occupational classification.
Your task is to accurately assign a job posting to the most suitable
NOCC 2022 minor class code. Please follow the instructions below.

The job responsibilities are:
{extracted_responsibilities}

Here is a list of potentially relevant NOCC minor classes
retrieved based on the job responsibilities:
{retrieved_nocc_candidates_with_definitions}

Based on all the information provided, determine the most accurate
NOCC 2022 minor class for this job description. Provide only the NOCC
code as your answer.

Answer:
```

Example: Classifying a Software Engineer Role

- **Input:** "Web Frontend, 00028 – Web Frontend Development Engineer Job Responsibilities: 1. Responsible for the frontend development of web designs, using HTML/CSS/JavaScript for PC and APP projects, coordinating the work of interface designers and developers. 2. Communicate with UI designers on interaction and visual design issues, optimize frontend functional design, solve cross-browser compatibility problems, and meet multi-terminal development needs. 3. Develop JavaScript program modules, write common libraries and frameworks, with experience in web application team management; familiar with processes such as design, coding, testing, and deployment. 4. Design and develop the frontend presentation layer and its interaction with the backend, working with backend engineers to formulate technical implementation plans

and service interfaces. Job Requirements: 1. College degree or above, with more than 2 years of pure frontend development experience in the Internet industry. 2. Proficient in mainstream JS frameworks, familiar with Angular2, Vue, ReactJs, RequireJS, etc., as well as tools like SASS, LESS, and Webpack. 3. Proficient in H5 application construction and component encapsulation; familiar with ES6 and Typescript syntax; experienced in CSS/JavaScript performance optimization and solving cross-browser compatibility issues. 4. Skilled in using various script debuggers, DOM viewers for problem locating, and various tools to detect web service performance and pinpoint bottlenecks. 5. Backend development experience (e.g., Java, PHP, Node.js) is a plus. 6. Strong team spirit, excellent learning and innovation ability, and capable of planning team training and development. Job Function Category: Web Frontend Development"

- **Extracted Key Responsibilities:** "Job Title: Web Frontend Development Engineer. Job Description: Responsible for frontend development of web designs, using HTML/CSS/JavaScript to build PC and APP projects, while coordinating the work between interface designers and developers."

- **Output:** `4-04-05(GBM 40405)` (Software and IT Service Professionals)

Application 2: Skill Classification

For this application, the objective is to identify and map skill-related phrases within each job posting to the most appropriate minor class in the Burning Glass skills taxonomy. The knowledge base for this task is the Burning Glass skills taxonomy, which contains over 442 minor classes.

Prompt Template

The following simplified template illustrates the prompt used for classifying each extracted skill phrase.

Prompt for extracting job skill descriptions

```
You are an expert in labor market skills analysis. Your task is to classify
     a given skill
phrase into the most relevant Burning Glass skill category (minor class).
    Please follow the instruction and finish the task below:


The job reuiqurements are:
{raw_job_description_text}


Please extract the skills needed to fullfill the job requirements:
```

Prompt for skill classification

```
You are an expert in labor market skills analysis. Your task is to classify
     a given skill phrase into the most relevant Burning Glass skill
    category (minor class). Please follow the instruction and finish the
    task below:

The skill required for a specific job is:
{extracted skill}

Here is a list of potentially relevant Burning Glass minor classes. Please
    choose the single best fit for the skill above:
{retrieved_burning_glass_candidates_with_definitions}

Based on the skill phrase and its context, determine the most accurate
    Burning Glass minor class. Provide only the category name as your
    answer.

Answer:
```

Example: Classifying a Software Engineers' Skills

- **Input:** "Job Requirements: 1. College degree or above, with more than 2 years of pure frontend development experience in the Internet industry. 2. Proficient in mainstream JS frameworks, familiar with Angular2, Vue, ReactJs, RequireJS, etc., as well as tools like SASS, LESS, and Webpack. 3. Proficient in H5 application construction and component encapsulation; familiar with ES6 and Typescript syntax; experienced in CSS/JavaScript performance optimization and solving cross-browser compatibility issues. 4. Skilled in using various script debuggers, DOM viewers for problem locating, and various tools to detect web service performance and pinpoint bottlenecks. 5. Backend development experience (e.g., Java, PHP, Node.js) is a plus.

6. Strong team spirit, excellent learning and innovation ability, and capable of planning team training and development."

- **Extracted Key Information:** "JavaScript , Angular2 , Vue , ReactJs , RequireJS , SASS , LESS , Webpack , H5 (HTML5) , ES6 , TypeScript , CSS/JavaScript Performance Optimization , Cross-Browser Compatibility , Script Debuggers , DOM Viewers , Web Service Performance Testing , Backend Development , Java , PHP , Node.js , Teamwork , Learning Ability , Innovation Ability , Team Training and Development Planning"

- **Output (Burning Glass Skill Minor Class):** "Collaborative Software, Middleware, Extensible Languages and XML, Integrated Development Environment (IDE), Business Strategy, Other Programming Languages, Scripting Languages, Software Development, Mobile Development, Software Development Tools, JavaScript and jQuery, Web Services, User Interface and User Experience (UI/UX) Design, Scripting, Social Skills, Web Design and Development, Employee Training, Java"

Appendix B. Appendix Results: External Shocks

Appendix A. COVID-19 Pandemic

We first estimate the baseline COVID specification:

$$COV_{t,\tau} = \alpha_0 + \beta_{RB}(t - t_{0,\tau}) + \beta_{COVID}\, I_t^{COVID} + \beta_{COVID,RB}\, I_t^{COVID} \cdot (t - t_{0,\tau}) + \delta_\tau + \varepsilon_{t,\tau}, \quad \text{(B1)}$$

and analogously with $Top5_{t,\tau}$ as the dependent variable, where $I_t^{COVID} = 1$ for quarters after 2020Q1. To further investigate heterogeneous effects across skill levels, we extend the specification to:

$$\begin{aligned} COV_{t,\tau} = \alpha_0 \; & + \beta_{RB}(t - t_{0,\tau}) + \beta_{COVID} \cdot I_t^{COVID} + \beta_{COVID,RB} \cdot I_t^{COVID} \cdot (t - t_{0,\tau}) \\ & + \beta_{lowskill} \cdot I_{\tau,t}^{lowskill} + \beta_{RB,lowskill} \cdot I_{\tau,t}^{lowskill} \cdot (t - t_{0,\tau}) \\ & + \beta_{COVID,lowskill} \cdot I_t^{COVID} \cdot I_{\tau,t}^{lowskill} \\ & + \beta_{COVID,RB,lowskill} \cdot I_t^{COVID} \cdot I_{\tau,t}^{lowskill} \cdot (t - t_{0,\tau}) \\ & + \delta_\tau + \varepsilon_{\tau,t}, \end{aligned} \quad \text{(B2)}$$

where $I_{\tau,t}^{lowskill} = 1$ if postings are classified as low-skill.

Appendix B. U.S. Sanctions

We collect data on all Chinese entities appearing on the BIS Entity List since its introduction in 1997. After excluding individuals, we identify 297 unique Chinese entities, including corporations, universities, research institutes, and government agencies. For each technology τ, we define the *first entity date* t_τ^{first} as the first quarter in which a sanctioned entity becomes one of the top five patent-producing institutions for that technology. This date marks the point when sanctions become directly relevant for the diffusion of that technology.

We then construct the indicator

$$I_{t,\tau}^{Entity} = \mathbb{1}\{\, t \geq t_\tau^{first} \,\},$$

which equals one if technology τ is affected by sanctions in quarter t.

We estimate

$$COV_{t,\tau} \;=\; \alpha_0 + \beta_{RB}(t - t_{0,\tau}) + \beta_{Entity}\, I_{t,\tau}^{Entity} + \beta_{Entity,RB}\, I_{t,\tau}^{Entity} \cdot (t - t_{0,\tau}) + \delta_\tau + \varepsilon_{t,\tau}, \quad \text{(B3)}$$

and analogously with $Top5_{t,\tau}$ as the dependent variable.