# The Content of Work Across Countries: An Open-Source Database of Job Postings

Fernando Saltiel*

In recent years, economists have drawn on job postings to examine changes in job requirements and the tasks employers ask workers to perform (Hershbein and Kahn, 2018, Deming and Kahn, 2018, Atalay et al., 2020). These postings provide detailed information on job vacancies, and comparisons with official statistics help economists relate the patterns they reveal to broader labor demand (Dalton et al., 2026). At the same time, the descriptions reveal how firms adopt new technologies and how demand for the associated skills spreads across employers and places (Acemoglu et al., 2022, Kalyani et al., 2025). In this context, job postings offer a way to examine how technological change affects the work employers demand and the expertise they seek, including the opportunities available to workers with less experience (Autor et al., 2024, Autor and Thompson, 2025).

In this project, I construct an open-source database of job postings from Common Crawl, a publicly available archive of the web. The database covers 76 countries and sixteen languages between 2018 and 2026. I link descriptions of advertised work to harmonized occupations, employers and locations, allowing researchers to distinguish changes in the occupations firms hire from changes in the tasks they request within those occupations. I use BERT-based text classifiers to identify the tasks described in these postings, drawing on recent advances in the analysis of text in economics (Ash and Hansen, 2023). This allows me to recover responsibilities that existing occupational classifications do not capture, such as work involving a newly introduced technology, and connect them to the qualifications employers request.

By drawing on publicly available sources, I aim to make both these measures and their construction methods available to other researchers. This gives researchers the flexibility to recover information that speaks directly to their questions, while using common classifications to compare the jobs in which it appears. For instance, a researcher studying entry into new work can distinguish experience with a particular technology from broader professional experience, or a degree requirement from an employer's willingness to accept equivalent experience. These distinctions matter for assessing which workers have access to emerging jobs and how employers value different combinations of skills and expertise (Deming and Kahn, 2018, Autor and Thompson, 2025).

I illustrate these opportunities by examining the work employers advertise as artificial intelligence spreads. I distinguish jobs that involve developing AI from those that involve using or overseeing it, and link these responsibilities to the skills and experience employers request. In doing so, I examine how employers' demand for expertise differs between specialized AI jobs and occupations where workers use the technology. The recent data reveal rapid growth in tasks that involve working with AI, alongside higher posted pay and more frequent experience requirements in jobs involving generative AI. These patterns illustrate how the database connects changes in the content of work to the qualifications employers seek.

*McGill University. Email: **fernando.saltiel@mcgill.ca**

## 1. Data and Measurement

To create the dataset, I use Common Crawl, a publicly available archive of web pages collected at regular monthly intervals (Common Crawl Foundation, 2026). In particular, I take advantage of the growing use of schema.org job-posting records, which report the title, employer, location and description of each job (Brinkmann et al., 2023). I recover these records from August 2018 through August 2026, allowing me to follow advertised labor demand over an eight-year period and add new observations as the archive expands.

I therefore use the job postings available in each monthly crawl to construct a sample of advertised labor demand. To avoid counting multiple copies of the same advertisement, I keep one observation per posting in each crawl. When a posting appears in a subsequent crawl, I observe it again and can follow how long it remains in the archive. This sample includes 183.4 million posting observations across 76 countries between 2018 and 2026. Developed countries account for most of the sample, including over 70 million observations for the United States and substantial coverage of Germany and the United Kingdom. Nonetheless, I also recover postings across the development spectrum, including India and Brazil, among others.

While the job titles in the sample describe the positions employers seek to fill, they are not necessarily comparable across contexts. As such, I clean and standardize these titles and map them into harmonized four-digit occupations using ISCO-08. I use these classifications to compare the prevalence of different occupations across countries and over time, while drawing on the descriptions to examine how work differs within the same occupation (Lo Bello et al., 2019, Atalay et al., 2024).

I additionally use information on the companies that post these jobs to identify employers and the locations where they seek to hire. Specifically, I standardize these companies' names and distinguish direct employers from staffing agencies and job boards. I then combine the employer information with the locations of the postings to follow each firm's labor demand across local labor markets. These links allow me to examine whether employers seeking workers for new technologies also change the other occupations and skills they demand in the same locations (Acemoglu et al., 2022).

More broadly, the employer information allows me to examine how multinational firms organize their labor demand across countries. I identify over 18,000 employers advertising at least ten jobs in each of at least two countries, providing scope to compare the responsibilities and qualifications the same firm requests in different labor markets. For instance, a firm may recruit workers to develop a technology in one country while seeking workers to apply it in another. I use these comparisons to examine whether technology-related work spreads across a firm's hiring locations and how its demand for expertise differs across them (Kalyani et al., 2025, Bastos et al., 2026). By linking these employers to information on their innovation activity, researchers can also examine how the work firms advertise relates to the technologies they develop.

Importantly, I draw on the raw job descriptions to uncover the tasks workers are expected to perform (Atalay et al., 2020). I use BERT-based text classifiers to identify task statements (Ash and Hansen, 2023), and plan to relate them to work activities described in O*NET and other sources. These comparisons build on my work with Markus Poschke and coauthors, which shows that differences in tasks within occupations help explain why fewer jobs can be performed remotely in developing countries (Gottlieb et al., 2021). The postings allow me to examine how such differences evolve as employers adopt new technologies, complementing evidence on technology and the task content of jobs across the development spectrum (Caunedo et al., 2023).

I further draw on the descriptions to recover the education, experience and credentials employers request, and am extending these measures to skills using Lightcast's skills taxonomy (Lightcast, 2026). I distinguish formal qualifications from requests for experience performing a particular activity or working with a specific technology. This allows me to examine whether employers seeking different work also seek different qualifications, even within the same occupation (Deming and Kahn, 2018). Moreover, the descriptions provide information on the alternatives employers accept, such as professional experience in place of a degree, which can help characterize the routes through which workers qualify for these jobs.

Lastly, I recover salary information and standardize the reported pay periods to compare compensation across postings. Posted pay provides a direct connection between the work employers advertise and the compensation they offer, complementing the information on tasks and qualifications (Batra et al., 2023). I use these measures to compare the pay associated with different responsibilities within the same employer and occupation, including jobs that combine established professional skills with knowledge of a new technology.

Finally, to understand aggregate patterns in labor demand, I weight the US postings in the proposed analysis to match the industry and state distributions of job openings in JOLTS, drawing on recent work comparing online postings with official vacancies (Dalton et al., 2026). These comparisons allow me to account for changes in labor demand across sectors and locations when examining the tasks and qualifications employers seek. I plan to draw on similar vacancy statistics in other countries to assess their coverage and construct comparable weights.

## 2. Artificial Intelligence and the Content of Work

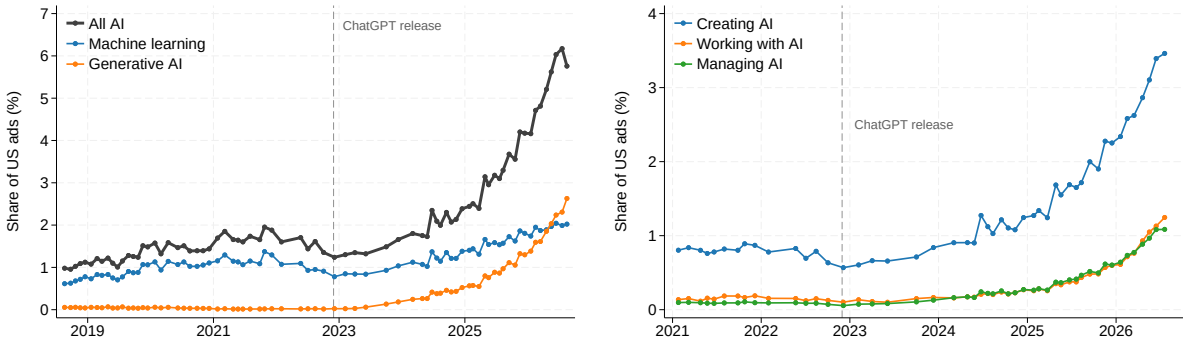
As noted above, I can use the raw descriptions to uncover the different ways in which employers incorporate AI into their work. For instance, a posting that mentions AI may ask workers to build a model, apply an existing tool, or oversee its use. As a result, I separate these components to examine how labor demand involving AI reaches a broad range of occupations and may affect employment within them. Moreover, I use the descriptions to uncover the type of AI involved, as jobs may draw on machine learning, natural language processing or generative AI. I connect these activities and technologies to the skills employers require or prefer, distinguishing demand for expertise in developing AI from experience using it and knowledge of the field in which workers apply it.

Figure 1 shows how AI demand has changed across technologies and activities in US English-language postings. Generative AI was rarely mentioned before the release of ChatGPT, but subsequently became an increasingly common part of advertised demand. The share of postings describing tasks that involve working with AI roughly doubled, from 0.6 percent in January 2026 to 1.2 percent just six months later. Employers thus increasingly seek workers who apply AI in their jobs, alongside those who develop the technology or manage its use.

I further examine how these responsibilities relate to the expertise employers seek. For example, the archive includes jobs developing AI products for legal discovery and jobs using AI to prepare traffic briefings for executives. These applications raise different questions about the combination of technical knowledge and professional experience a worker needs. In preliminary comparisons of US postings from January through July 2026, I find that postings involving generative-AI work offer about 5 percent higher posted pay and more frequently list experience requirements than other postings within the same employer, four-digit occupation and crawl.[1] Among postings that specify years, the minimum experience requested is similar.

[1] I classify a posting as generative-AI work when it names generative AI and describes a task that involves creating or using AI.

**Figure 1:** AI Demand in US Job Postings




**(a)** By technology **(b)** By AI activity

*Notes:* Figure 1 shows AI demand among classified US English-language postings with at least ten description words. Panel (a) reports demand by technology through August 2026. Panel (b) distinguishes tasks involving the creation, use and management of AI, from 2021 through July 2026. Creating AI includes developing models and systems; working with AI includes applying it to activities such as analysis, coding and reporting; managing AI includes overseeing projects, teams and implementation. Each point reports the share of postings in the corresponding crawl, counting each posting once. Dashed lines mark the November 2022 release of ChatGPT.

The distinction between the type of experience employers seek and its duration gives me a way to examine how new tasks relate to the expertise they value (Autor and Thompson, 2025).

More generally, I use the dataset to examine how these forms of work spread across firms and countries. I follow multinational employers to compare the activities and qualifications they request across their hiring locations, and to ask whether AI development and use spread together. Researchers can draw on the same descriptions to study other technologies, tracing the responsibilities employers introduce and the qualifications they seek as those technologies spread.

## 3. Public Availability

Given the open-source nature of the project, I plan to make a significant share of the data publicly available. Following I3's emphasis on open data for research on innovation, I will share the code used to construct the measures, the occupational and geographic crosswalks, and the derived information on tasks, qualifications and AI-related work (Innovation Information Initiative, 2026). I also plan to provide aggregate data by country, period, occupation and employer, so that researchers can examine changes in labor demand without processing the entire archive. Alongside these files, I will describe the variables and provide examples that connect the measures to the underlying descriptions. Researchers can then adapt the definitions to their own questions and compare the resulting patterns using the same occupations, employers and locations. An existing public explorer already presents the AI-demand measures.[2]

I plan to submit the derived panels for inclusion in I3's data workspace and update them as Common Crawl releases new observations. By connecting employers to patents and scientific publications, researchers could examine how firms' innovation relates to the work and expertise they demand, complementing I3's existing resources (Marx and Fuegi, 2020, Marx and Shvadron, 2025). For instance, researchers could follow the tasks firms advertise around the introduction of a new technology and examine whether demand for the associated skills appears across their hiring locations. In this way, the project provides a basis for studying both the development of new technologies and the changes in work that accompany their use.

[2]The interactive explorer is available at https://ai-related-demand-explorer.fernando-saltiel.chatgpt.site/.

## References


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

Ash, E. and Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, 15:659–688.

Atalay, E., Phongthiengtham, P., Sotelo, S., and Tannenbaum, D. (2020). The evolution of work in the united states. *American Economic Journal: Applied Economics*, 12(2):1–34.

Atalay, E., Sotelo, S., and Tannenbaum, D. (2024). The geography of job tasks. *Journal of Labor Economics*, 42(4):979–1008.

Autor, D., Chin, C., Salomons, A., and Seegmiller, B. (2024). New frontiers: The origins and content of new work, 1940–2018. *The Quarterly Journal of Economics*, 139(3):1399–1465.

Autor, D. and Thompson, N. (2025). Expertise. *Journal of the European Economic Association*, 23(4):1203–1271.

Bastos, P., Stapleton, K., Taglioni, D., and Wei, H. Y. (2026). Firm networks and global technology diffusion. *Journal of International Economics*, 161:104240. Earlier version: World Bank Policy Research Working Paper, 2024, doi 10.1596/42154.

Batra, H., Michaud, A., and Mongey, S. (2023). Online job posts contain very little wage information. NBER Working Paper 31984, National Bureau of Economic Research.

Brinkmann, A., Primpeli, A., and Bizer, C. (2023). The web data commons schema.org data set series. In *Companion Proceedings of the ACM Web Conference 2023*, pages 136–139.

Caunedo, J., Keller, E., and Shin, Y. (2023). Technology and the task content of jobs across the development spectrum. *The World Bank Economic Review*, 37(3):479–493. Earlier version: NBER Working Paper 28681, 2021, doi 10.3386/w28681.

Common Crawl Foundation (2026). Common crawl: Overview. Accessed 2026-09-23.

Dalton, M., Kahn, L. B., and Mueller, A. I. (2026). Do online job postings capture job vacancies? NBER Working Paper 35697, National Bureau of Economic Research.

Deming, D. and Kahn, L. B. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, 36(S1):S337–S369.

Gottlieb, C., Grobovšek, J., Poschke, M., and Saltiel, F. (2021). Working from home in developing countries. *European Economic Review*, 133:103679.

Hershbein, B. and Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? evidence from vacancy postings. *American Economic Review*, 108(7):1737–1772.

Innovation Information Initiative (2026). Innovation information initiative (i3) governance document, version 1.0. Effective March 1, 2026. Accessed 2026-09-23.

Kalyani, A., Bloom, N., Carvalho, M., Hassan, T., Lerner, J., and Tahoun, A. (2025). The diffusion of new technologies. *The Quarterly Journal of Economics*, 140(2):1299–1365.

Lightcast (2026). Lightcast skills taxonomy (lightcast open skills). Web page, 34,000+ skills drawn from job postings and profiles. Accessed 2026-09-23.

Lo Bello, S., Sanchez Puerta, M. L., and Winkler, H. (2019). From Ghana to America: The skill content of jobs and economic development. Policy Research Working Paper 8758, World Bank, Washington, DC. Also circulated as IZA Discussion Paper 12259.

Marx, M. and Fuegi, A. (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal*, 41(9):1572–1594.

Marx, M. and Shvadron, D. (2025). The i3 BigQuery data workspace: Shared infrastructure for open science. NBER conference paper c15344, National Bureau of Economic Research.