Federal Reserve Bank of Dallas

Remote Work Across Jobs, Companies and Space

Stephen Hansen, Peter John Lambert, Nick Bloom, Steven J. Davis, Yabra Muvdi, Raffaella Sandun and Bledi Taska

Working Paper 2629 **August 2026**

Research Department

https://doi.org/10.24149/wp2629

Remote Work Across Jobs, Companies and Space[*]

Stephen Hansen[†], Peter John Lambert[‡], Nick Bloom[§], Steven J. Davis[±], Yabra Muvdi[°], Raffaella Sadun[¶] and Bledi Taska[¤]

August 11, 2026

Abstract

The pandemic catalyzed an enduring shift to remote work. To measure this shift, we develop a large language model (LLM), fine tune and assess it using 30,000 human classifications, and apply it to nearly 600 million job vacancy postings across five English-speaking countries. Our model achieves a 98% classification accuracy, greatly outperforming dictionary-based approaches, and matching or surpassing the performance of frontier AI models at a fraction of the cost. From 2019 to 2026, the share of postings that indicate new employees can work remotely at least one day per week rose more than three-fold in the U.S. and by a factor of five or more in Australia, Canada, New Zealand, and the U.K. These developments are highly non-uniform across and within cities, industries, occupations, and companies. Even when focusing on employers in the same industry competing for talent in the same occupations, we find large differences in the share of job postings that explicitly offer remote work.

Keywords: remote work, hybrid work, work from home, job vacancies, text classifiers, large language models, pandemic impact, labour markets

JEL Codes: E24, O33, R3, M54, C55

---

[*]Thanks for outstanding research assistance to Miaomiao Zhang and Kelsey Shipman, who supported the data analysis. Hansen gratefully acknowledges financial support from ERC Consolidator Grant 864863, Lambert from the London School of Economics STICERD PhD research grant and the Commonwealth Scholarship Commission, Bloom from the Smith Richardson and John Templeton Foundations, Davis from the Templeton Foundation and the Booth School of Business at the University of Chicago, and Sadun from Harvard Business School. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System. Data accompanying this paper can be found at www.WFHmap.com.

[†]Stephen Hansen, University College London and Federal Reserve Bank of Dallas, stephen.hansen@ucl.ac.uk.

[‡]Peter John Lambert, London School of Economics and University of Warwick.

[§]Nick Bloom, Stanford University.

[±]Steven J. Davis, Hoover Institution.

[°]Yabra Muvdi, Digitec Galaxus AG.

[¶]Raffaella Sadun, Harvard Business School.

[¤]Bledi Taska, TalentNeuron.

1 Introduction

The COVID-19 pandemic propelled an enormous uptake in hybrid and fully remote work. Over time, it has become clear that this shift will endure long after the initial forcing event. In 2025, one-quarter of full workdays in the U.S. take place at home or other remote locations (Barrero et al., 2021). The pandemic also drove large, enduring increases in remote work in dozens of other countries (Criscuolo et al. 2021, Aksoy et al. 2022). There are few, if any, modern precedents for such an abrupt, large-scale shift in working arrangements.

Most efforts to quantify and characterize this shift rely on surveys of workers and employers, digital tracking of foot traffic, or remote-work feasibility assessments. We rely instead on information in job vacancy postings. Specifically, we consider the full text of nearly 600 million postings in five English-speaking countries. We count a posting as offering remote work when it states that a new employee can work remotely at least one day per week; so our measure pools hybrid and fully remote arrangements. To examine this text at scale, we build and deploy a large language model (LLM) that we fit, test, and refine using 30,000 classifications generated by human readings.

Vacancy postings pertain to the flow of new jobs rather than the stock. In addition, postings that promise remote work two days a week, for example, entail a commitment–or at least a statement of intent–that extends into the future. For both reasons, postings need not show the same pattern of remote work as employment. Indeed, the remote-work share of postings is smaller than the remote-work share of employment, especially in the early stages of the pandemic. And while the incidence of remote work among the employed fell markedly in the two years after spring 2020, we show that the remote-work share of postings rose sharply over the same period.

The share of postings that say new employees can work remotely one or more days per week was small before the pandemic: 0.8% to 1.5% in Australia, Canada, and New Zealand as of 2019, about 2% in the U.K., and about 3.5% in the U.S. From 2019 to 2025, this remote-work share rose more than three-fold in the U.S. and by a factor of five or more in the other countries. As of mid-2026, the remote-work share stands at about 9-10% in Australia, Canada, and the U.S., and nearly 17% in the U.K.—a "new normal" that shows no meaningful sign of reverting to pre-pandemic levels. New Zealand, a partial exception, shows some reversion from its 2023 peak while remaining well above pre-pandemic levels.

Remote-work posting shares vary greatly across occupations and cities. Looking across occupations, the remote-work share correlates positively with computer use, education, and earnings. Finance, Insurance, Information and Communications have especially high remote-work shares. Chicago, London, New York, San Francisco, Toronto, and other cities that

function as business service hubs have high remote-work shares. These differences have widened since the pandemic struck. In a linear regression, 61% of the variation across occupations in 2025 remote-work shares is accounted for by their 2019 shares. In contrast, just 13% of city-level variation in 2025 remote-work shares is accounted for by 2019 shares.

We also find that the shift to remote work is highly non-uniform across same-industry employers, even in the same occupational category. This emergent heterogeneity on the demand side expands opportunities to satisfy preferences over work arrangements on the supply side.[1] This non-uniformity result yields another important message: in many occupations, it is misleading to think of remote-work suitability as a purely technological constraint. Remote-work intensity is, instead, an outcome of choices about job design and how to operate an organization. These choices are influenced by the external environment and subject to shock-induced shifts. In line with this view, Aksoy et al. (2022) find that employers plan higher levels of work from home after the pandemic ends in countries that experienced longer and stricter government-mandated lockdowns during the pandemic.

Our approach to studying the remote-work phenomenon has several noteworthy strengths. First, our data cover almost all vacancies posted online by job boards, employer websites, and vacancy aggregators from January 2014 through June 2026 in our five countries. Coverage on this scale is infeasible with survey methods. Second, postings typically describe the job and its attributes in considerable detail, as suggested by a median posting length of 347 words. Comparable detail is hard to obtain from other sources, especially at scale.[2] Third, we develop a language-processing model that reads and classifies postings in an automated manner. Our model achieves a 98% accuracy rate in flagging jobs that allow for remote work, greatly outperforming dictionary methods. It also matches or surpasses the classification performance of frontier AI models at less than 1% of their processing cost. Fourth, we assess the external consistency of our vacancy-based measure against the American Community Survey and the Current Population Survey, finding strong correlations (0.70–0.93) across metropolitan areas and occupations.

The combination of scale, rich text data, granularity, and automation lets us track the shift to remote work across occupations, industries, locations and employers, making it possible to explore new questions about its extent, origins, and consequences. We post and

[1] On preference heterogeneity in regards to remote work, see Bloom et al. (2015), Mas and Pallais (2017), Wiswall and Zafar (2018), Barrero et al. (2021), and Aksoy et al. (2022).

[2] Previous research exploits the detail in vacancy postings to study technical change, the cyclicality of skill requirements, their relationship to wages, how compensation and other job attributes affect applicant flows and, of course, to classify jobs in a fine-grained manner. Examples include Modestino et al. (2016), Deming and Kahn (2018), Hershbein and Kahn (2018), Davis and Samaniego de la Parra (2020), Forsythe et al. (2020), Marinescu and Wolthoff (2020), and Acemoglu et al. (2022).

update many of our statistics at www.WFHmap.com, which draws 10,000 unique visitors a year. Our data have already informed many studies that speak to policy-relevant issues.[3] Examples include Autor et al. (2023) on post-pandemic wage compression, Barrios et al. (2024) on work from home (WFH) and entrepreneurial entry, Cowan and Garcia (2024) on how remote work shapes political preferences across U.S. counties, Bloom et al. (2026) on the implications of remote work for the employment of disabled persons, Lambert and Schindler (2026) on how WFH affects the career ladder, and Davis et al. (2026) on WFH and fertility. Our data are also featured in two Economic Reports of the President (Council of Economic Advisers, 2024; Council of Economic Advisers, 2025).

We build our classifier by fine-tuning a small, open-source language model on human labels. Fine-tuning a pretrained transformer in this way is common in the machine learning literature. In economics, it has been applied to corporate disclosures (Huang et al., 2023) and central bank communications (Pfeifer and Marohl, 2023), for example. The technique itself is well known, but the scale of our application and training sample has little precedent in the economics literature. We exploit that scale to quantify model performance as a function of training sample size and to assess how much training is needed to match the performance of frontier models. Our approach is fully open-source and replicable, a feature that is hard to achieve with proprietary models and their rapid release cycles.

We use DistilBERT, a smallish LLM developed by (Sanh et al., 2020) that descends from the BERT model introduced by Devlin et al. (2019).[4] First, we pre-train DistilBERT on 900,000 sequences drawn from vacancy postings. This step familiarizes DistilBERT with the (heterogeneous) structure of job ads. Second, we consider 10,000 text sequences drawn from vacancy postings and assign three human readers to label each one. The reader assigns a positive label if the text says the job allows work from home or other remote location one or more days per week. Third, we fit the pre-trained DistilBERT framework to the human labels to obtain a model that classifies each posting. Finally, we apply the resulting "Remote Work in (Job) Openings" (RWO) language model to classify every posting in our corpus.

The RWO model greatly outperforms a dictionary used to classify the remote-work status of vacancy postings in earlier work.[5] Expressions like "work from home care facilities" and "requires a Home Office work permit" suggest the difficulties that arise when applying dictionary methods, which in job ads yield high classification error rates that vary greatly

[3]We first released our code and accompanying data in March 2023, a few months after the debut of ChatGPT.

[4]BERT and DistilBERT are pre-trained on the full English-language Wikipedia corpus and the Toronto Book Corpus. For a helpful non-technical overview of BERT.

[5]Previous work that uses dictionaries to measure remote work in job postings includes Adrjan et al. (2021), Bai et al. (2021), Bamieh and Ziegler (2022), Draca et al. (2022), and Alipour et al. (2023).

over time and across occupations. Logistic regressions and generic AI models offer large improvements over dictionary methods. Our RWO classifications offer further improvements, outperforming other methods as measured by accuracy rates, precision, and F1 scores. [6] This result is important in light of the increasing reliance on generic AI models to implement zero-shot and few-shot approaches to classification tasks.[7]

A prominent line of research classifies occupations as suitable or unsuitable for remote work based on descriptions of work activities and experiences.[8] Our analysis highlights some limitations of this approach. First, remote-work intensity is a malleable feature of jobs, occupations, and organizations. Second, classifications based on suitability assessments explain little of the variation in remote-work posting shares. For occupations that Dingel and Neiman (2020) classify as unsuitable to be done entirely from home, the remote-work share of U.S. postings in 2025 ranges from 0 to 38% with a mean of 4% and standard deviation of 6%. For occupations they classify as suitable for work from home, the share ranges from 0.05% to 59% with a mean of 16% and standard deviation of 11%.

Another prominent line of research surveys workers and employers to study working arrangements. Barrero et al. (2020), Bartik et al. (2020), Bick et al. (2023) and Brynjolfsson et al. (2020) document and characterize the enormous uptake in work from home in spring 2020. Bartik et al. (2020), Barrero et al. (2021) and Ozimek (2020) use employer plans and other forward-looking survey data to forecast that the big shift to remote work will endure. Relative to our approach, the survey-based approach is more useful for eliciting the perceptions, attitudes, and expectations of workers and employers. Our approach offers several other distinct advantages, as discussed above.

The next section describes our vacancy posting data and develops our classification model. Section 3 assesses the model's performance in absolute terms, and relative to other approaches. Section 4 sets forth our main findings related to remote-work intensity over time and across countries, cities, occupations, and more. We also compare our remote-working posting shares to survey-based measures of remote work. Section 5 concludes.

[6] Shapiro et al., 2022 develops a BERT-based model and finds little gain relative to dictionaries in detecting news sentiment. They train on 800 human-labeled articles. Consistent with our evidence in Figure 1, gains from fine-tuning at that scale are modest. Bajari et al. (2021) and Bana (2022) use BERT to predict prices from Amazon product reviews and wages from job posting text, respectively. Each paper achieves high predictive performance. Their applications don't involve the use of human-generated labels.

[7] See Ash and Hansen (2023) and Ash et al. (2025) for recent reviews.

[8] Dingel and Neiman (2020) is the most influential example. Other examples include Rio-Chanona et al. (2020), Mongey et al. (2021), and Adams-Prassl et al. (2022). Like us, Adams-Prassl et al. (2022) concludes that remote-work intensity varies greatly across jobs within occupations.

2 Data and Measurement

To measure remote-work posting shares, we exploit a near-universe of online job postings from January 2014 through June 2026 for our five countries.

We extract 10,000 text sequences from selected postings and ask humans to read them. Each sequence is about 45 words long, and the average posting has about six sequences. Breaking postings into sequences facilitates human and algorithmic classification at scale, as we discuss below. Our human readers answer this question: 'Does this text explicitly offer an employee the right to remote-work one or more days a week?', yielding a binary classification. The pairwise agreement rate between readers exceeds 90 percent.

We use DistilBERT, a derivative of BERT, as the foundation model for our remote work classifier.[9] DistilBERT is trained to optimize the same loss function as BERT but with many fewer parameters—it has 66 million parameters as opposed to 110 million in the original model. It does so by using BERT as a 'teacher' that guides training (Hinton et al., 2015). We modify off-the-shelf DistilBERT in two ways. First, we update its weights to resolve word-prediction tasks on a sample of vacancy posting text, whose context might differ from generic English (*unsupervised fine-tuning*). Second, we use the human labels to further update the model for the remote work classification task (*supervised fine-tuning*). We call this the "Remote Work in (Job) Openings" (RWO) model. We will show that RWO achieves near-human performance in its classification task, and that it outperforms a variety of other approaches. We describe our approach in some detail, because we think it has useful applications to many other text-analysis tasks in economics and other fields.

2.1 Job Vacancy Data

We examine online vacancy postings collected by Lightcast (formerly Emsi Burning Glass), an employment analytics and labor market information firm. Lightcast scrapes postings from more than fifty thousand online sources that include vacancy aggregators, government job boards, and employer websites. Lightcast claims to cover a "near-universe" of online postings in our five countries during the period covered by our analysis. We drop a small fraction of postings through quality filters on posting sources and daily posting volumes. See Appendix A for a description of our data and pre-processing steps.

[9]BERT stands for Bidirectional Encoder Representations from Transformers. Transformers are a deep-learning method in which every output element is connected to every input element of a text sequence. This allows the meaning of a particular word to depend on the context of surrounding words, which as we show below is crucial in our setting. See Phuong and Hutter (2022) for a formal overview of how Transformers work. Vaswani et al. (2017) is the seminal contribution.

Burke et al. (2020) compare vacancy postings in Lightcast data for the United States to job vacancy data from the U.S. Job Openings and Labor Turnover Survey (JOLTS). The two sources are reasonably well aligned, but the JOLTS data show larger vacancy shares in food services, public administration and construction and smaller shares in finance, insurance, healthcare, social assistance and educational services.

For each vacancy posting in our dataset, we have access to a plain text document scraped from the job listing. We also observe the posting date, employer name, occupation, employer location, industry, and more. We consider postings listed from January 2014 through June 2026, dropping those with an unknown occupation (less than 1%). We use a 5% random sample of postings before January 2019, and the universe of postings thereafter. The resulting dataset covers nearly 600 million online vacancy postings in five countries, spanning 13.4 million employers and more than 78 thousand cities. Table 1 provides more information.

For our baseline results, we re-weight the postings in each country-month cell to match the U.S. occupational distribution of new postings in 2024. We do so because raw posting shares confound differences in the propensity to offer remote work with differences in the mix of jobs being advertised, which varies across and within countries and over time. Holding the occupational distribution fixed at a common benchmark removes compositional variation, so that a gap between two countries, or a movement within one country, reflects how often employers offer remote work within occupations rather than which occupations happen to be hiring. In practice, this gap is small. For example, the gap in the 2025 share of job postings offering remote/hybrid between the US and UK falls from roughly 6.6 percentage points in the raw data to 5.0 percentage points once both countries are placed on the same occupational distribution. Appendix B reports the country series under alternative weighting schemes: Table B.4 gives each country's share under these baseline weights and under its own 2019 occupational composition, and Figure B.2 plots the raw, unweighted series.

2.2 The Measurement Problem

The measurement problem we face is to determine whether each job posting allows a new hire to work remotely, understood here to encompass both fully remote and hybrid positions. We adopt a binary classification approach, and refer to a 'positive' posting as one that mentions the ability to work remotely, and a 'negative' posting as one that does not. For positions that offer hybrid working arrangements, we use a threshold of at least one day per week for our positive classification. This approach effectively measures an employer's willingness to commit *ex ante* to offering flexibility in work location. Negative postings may in fact be associated with work-from-home positions, for example because the ability to work from

home is assumed by market participants to be feasible in particular jobs, or because the employer prefers to bargain over work arrangements during the hiring process rather than make a prior commitment. We return below to discuss the interpretation of our measure, and first focus on developing an accurate and robust classification.

The most precise way of classifying postings is arguably via direct human reading. Given the volume of our raw text data, however, this approach is not feasible to scale and an automated classification is essential. A traditional approach in the text-as-data literature uses a dictionary of keywords whose presence is treated as a positive classification. As an initial step, we use the keywords in Table C.1 to classify job postings as positive or negative. While we do not claim the dictionary of terms is fully optimized, it is in line with others in the literature for classifying postings as work-from-home (Adrjan et al., 2021).

Job postings classified by keywords yields notable errors, as illustrated in Table 2. False positives include references to company home offices and working at someone's home to provide health care. Another, more concerning, issue involves shifts in the structure of job ads during COVID-19 that correlate with the incidence of false positives. After early 2020, many postings began featuring new text fields along the lines of "work from home allowed?" A naive dictionary method could then infer that a posting allows WFH even when the rest of that field reads "no." Extending dictionaries to include negation mitigates this problem, but we show in section 3 that our approach is much more accurate. Table 2 also lists examples of false negatives, which illustrate the many and complex ways that ads refer to remote work. Treating this linguistic variety with a fixed set of keywords is a major challenge.

2.3 Our Approach to Classification

Our approach has three steps. First, human auditors read and classify 10,000 pieces of text extracted from job ads, yielding 30,000 labels. Second, we train DistilBERT using these human classifications. Third, we take this predictive model out-of-sample to classify each job ad as either positive or negative. The intent is to scale the accuracy of human readings to the entire dataset. We refer to the resulting classifier as the "Remote Work in (Job) Openings" (RWO) LLM, or simply RWO. We now describe our approach more fully, placing details in Appendix C.

2.3.1 Breaking Up Job-Ad Text into Sequences

While we ultimately wish to classify job postings, we initially label and classify smaller units of text we refer to as *sequences*. The first reason for doing so is that human labeling of entire

job postings is prone to a high error rate because of their length and complexity. The second reason is that the typical posting has a great deal of information unrelated to remote work, for example descriptions of skills required for the job, tasks involved, etc. Mixing text relevant to WFH with a great deal of irrelevant text introduces noise into the classification algorithm. Splitting each posting into its title, structured fields and paragraph-level sequences keeps the unit of classification short enough for human readers to label reliably and for a smallish language model to learn precisely.

Our procedure for generating sequences has three salient features. First, postings always begin with a job title, e.g. "Software Programmer familiar with R and Python", which we extract as a single sequence. Second, most postings also contain distinct bullet points and structured fields, each of which we treat as a single sequence, unless it exceeds a length threshold. In that case, we treat it as multiple sequences. Finally, the rest of a posting typically contains standard prose in one or more paragraphs. We treat each paragraph as a single sequence, unless it passes a length threshold, in which case we break it into multiple sequences. This procedure yields 3.4 billion sequences from nearly 600 million job postings.

2.3.2 Human Labels for Training and Evaluation

We chose 10,000 sequences for manual labeling. We selected one quarter at random from the set of sequences that contained a set of dictionary terms listed in Table C.1. Another quarter was chosen to contain a broad set of terms that might reflect WFH language, including the generic terms 'remote', 'home', 'work', 'location'; any word that begins with 'tele'; and any two-word sequence that begins with 'remote'. Another quarter consisted of sequences that might confound a classifier, including 'home repairs', 'nursing home', 'remote construction', etc. The final quarter was a random sample of sequences not satisfying the three aforementioned criteria. Each portion of the label sample is balanced across quarters from 2014Q1 through 2021Q3. We also balance the sample evenly across countries, drawing one quarter from each of the USA, UK, Canada, and a further one quarter from the pooled Australia and New Zealand data to account for varying English idioms in different geographic locations.

We tapped Amazon Mechanical Turk to generate labels. To ensure high-quality auditors, we set up an initial screening test that required prospective auditors to label 20 sequences that we had personally classified. Only auditors that made at most one error were allowed to label the full set. Another quality control strategy was to pay around 25% above typical market rates for labeling tasks. This motivated auditors who passed initial screening to continue on the project. Several wrote to us seeking clarification on ambiguous cases, which went beyond what the platform required of them.

Each of the 10,000 sequences is labeled by three distinct auditors. For 66.9% of the sequences, all three auditors assign a negative label. For another 25.5%, all three assign a positive label. The auditors disagree on the remaining 7.6%. While we chose half of the sequences to contain a word with the potential to denote WFH, only 29.2% of them are labeled as positive by two or three auditors.

2.3.3 Developing the RWO model

The field of language processing has been revolutionized by models that use context to interpret meaning. Consider two sentences: 'Some of the deep-sea wells we operate are in remote locations'; and 'We are pleased to offer opportunities for remote work'. Each includes 'remote' but only the second indicates remote work in our sense. The important point is that the interaction of 'remote' with surrounding context words determines the meaning of these sentences. Moreover, not all context words are equally informative: for example, in the first sentence 'deep-sea', 'wells', and 'locations' are more important than 'some' and 'we' in understanding the meaning of 'remote'. *Self-attention* (Vaswani et al., 2017) is a mathematical construct that allows vector representations of individual words to interact with each other to form new vectors that encode the meaning of sequences. These interaction weights effectively determine which words should be "paid attention to" in resolving these meanings. Self-attention is the key idea that powers all widely-used LLMs. DistilBERT is one such model. See Ash and Hansen (2023) and Ash et al. (2025) for further details.

We make two main modifications to the off-the-shelf DistilBERT model to build RWO. First, the initial DistilBERT parameters are obtained by predicting randomly deleted words in generic English from surrounding context words. We instead update these parameters to predict randomly deleted words in a sample of 900,000 job posting sequences that is balanced across all years and countries. This step creates word representations that are specific to the language of job postings.

Second, we further modify off-the-shelf DistilBERT to predict human labels from vector representations of job posting sequences. We split our labeled sequences into training and test sets of 5,950 and 4,050 sequences, respectively. The prediction problem is conducted at the label rather than sequence level, so there are 3 * 5,950 = 17,850 total observations in the core training sample of labeled data.[10] Appendix C details how we specify the prediction model's hyper-parameters. Table 3 provides an illustration of which words are influential

[10]During an initial exploratory phase, we labeled a sample of around 10,000 additional sentences (rather than sequences) using a combination of Mechanical Turk, hired research assistants, and ourselves. Since these are also potentially informative, we include them in the training set. In most cases, these sentences only received a single label and so in total generate 11,574 additional labels in the training set.

in the classification problem, and compares our RWO approach to a dictionary approach. Crucially, the weights attached to words are learned by the algorithm rather than imposed *ex ante* by researchers, and the weight on a particular word depends on the surrounding words. Below, we compare the test-set accuracy of our model with that of other algorithms.

2.3.4 Predicting Remote Work Language at Scale

Finally, we use the prediction model to assign a continuous probability to all sequences in our corpus. The higher the probability, the more confidence the model has that this sequence denotes an offer of remote work arrangements. Figure B.1 plots a histogram of the share of sequences that fall in different probability intervals. The distribution is bimodal at the lowest and highest probability bins, with the former dominating the distribution. As expected, most sequences do not contain WFH language because, as we show below, most job postings do not explicitly mention the possibility to WFH and, among those that do, the majority of sequences discuss other features of the position. The bimodality of the distribution shows that the classification algorithm usually produces a clear prediction, in line with the high agreement rates among human auditors. We use an 0.5 threshold for treating a sequence as positive for WFH, but it's clear that our results are not sensitive to this particular cutoff.

2.3.5 Aggregating Measurement back to Job Postings

We have conducted all the analysis so far at the sequence level, but are ultimately interested in a posting-level classification. For this, we use a simple 'max' rule and positively classify a job posting if it contains one or more positive sequences. Table B.1 shows the number of positively classified sequences in each job ad. We can see that among the positive job ads (those with one or more positive sequences), the majority have just a single positive sequence. This reduces concern that the algorithm produces correlated false positive hits at the posting level.[11] This posting-level classification constitutes the final output from our RWO model, which we use to study the adoption of remote work.

2.3.6 Public Access to RWO

So that researchers can interact with and study the properties of our model, we make available a simple online tool that allows one to input arbitrary text and receive a predicted probability as output. The URL is `https://huggingface.co/spaces/yabramuvdi/wfh-app-v2`,

[11] We manually read a number of randomly drawn postings with more than five positive sequences and found no instance of the algorithm failing.

which will reproduce the same probabilities as in the paper. Users who find biases or other weaknesses in the predictions are welcome to contact the authors with their findings, which can be incorporated into future work.

2.3.7 Computational Performance of RWO

One constraint on implementing large-scale NLP models is computational. To provide some performance guidelines, Table B.2 tabulates the hardware we use for each training stage, the time taken, and the cost involved. All estimation is done on the Google Cloud Platform. In neither time nor money terms is the implementation of RWO particularly costly: the two training stages together take 15 hours and cost $45, and scoring our full corpus of 3.4 billion text sequences brings the all-in total to roughly $9,600 at current prices ($2.80 per million sequences). Our view is that researchers should therefore not view computational costs as a major impediment to customized, purpose-built LLMs. On request, we make available all code to efficiently train RWO and apply it out-of-sample. Interested researchers should register interest at WFHmap.com.

3 Assessing the Performance of RWO

Above we highlight instances in which the presence or absence of keywords is insufficient to correctly classify a selection of job posting texts due to the complexity of surrounding context. In order to quantify the gains from adopting our approach, we now compare classification performance across various algorithms. To do so, we adopt a standard approach in the machine learning literature and randomly split the 10,000 human-labeled sequences into training and test sets (of sizes 5,950 and 4,050, respectively). We then train RWO just on the training data and use the fitted model to assign a predicted value to each test-set observation. We compare the following methods, with full details in Appendix C:

1. *All Zero.* Each test-set observation is assigned a 0 to match the modal outcome.

2. *Dictionary.* We use a term set similar to that from Adrjan et al. (2021),[12] and count an observation as positive if it contains a term from this set.

3. *Dictionary with Negation.* Shapiro et al. (2022) show that accounting for negation can improve the performance of dictionaries. We adopt a similar method, count a dictionary term as indicating remote work only if not negated by nearby text.

[12]The terms are reported in Table C.1. The remote work measures in Adrjan et al. (2021) are based on data from Indeed which potentially has a different structure from the Lightcast data.

4. *Logistic Regression.* Adams-Prassl et al. (2020) use Lightcast postings for the UK to measure the prevalence of flexible work schedules. They manually annotate 7,000 texts and fit a (penalized) logistic regression model for classification. We implement a similar logistic model on our training data and use it to classify test data.

5. *Logistic Regression with Negation.* We expand the feature set of the logistic regression to incorporate negation and re-estimate it on the training data.

6. *Zero-shot learning with generic AI models.* We use four distinct generic AI models to classify sequences with zero-shot learning: two non-reasoning models (GPT-4o and Claude Sonnet 4.6) and two reasoning models (GPT-5.5 and Claude Opus 4.8).

7. *RWO with Generic English.* Here we only update DistilBERT via supervised fine-tuning and skip the unsupervised fine-tuning step.

Table 4 reports the test-set performance for all methods, on the audit sample in columns (1)–(4) and reweighted to the population composition of postings in columns (5)–(8). The reweighted panel is the economically relevant comparison, and we return to it below. A straightforward metric is the error rate, i.e. the fraction of mis-classified texts. On this measure, RWO achieves the lowest error rate of any method we test, 0.02, on a par with the strongest frontier models and far below the dictionary and bag-of-words alternatives.[13] The nearest competitors are GPT-5.5 and Claude Sonnet 4.6, which also achieve error rates of 0.02, followed by GPT-4o, Claude Opus 4.8, and RWO without unsupervised fine-tuning (error rates of 0.03), while the dictionary method's error rate is six times higher.

A more standard performance metric in the machine learning literature is the F_1 score, which accounts for a classifier's ability to recover true positives (*recall*) and the share of predicted positives that are true positives (*precision*).[14] The F_1 score ranges from 0 and 1, where higher values indicate better performance. RWO again attains the best score, with the leading frontier models close behind.

A potential concern is that the distribution of positive and negative postings in the test data does not correspond to that of the full population: the data extracted for labeling is designed to over-represent positive cases. To assess classification accuracy in the population, we re-weight the test-set metrics to the base rate in the corpus itself. Applying RWO to

[13] This error rate is consistent across countries and years. When broken down by country, the test set error rate is 0.02 in each case. When broken down by year, the test set error rate is 0.02 in each year except for 2015 (error rate 0.03) and 2014 (error rate 0.01).

[14] Let TP, FP, TN, and FN be the number of true positives, false positives, true negatives, and false negatives, respectively. Precision is TP / (TP + FP), and recall is TP / (TP + FN).

every sequence in our data, 2.27% are classified as positive over January 2022 through June 2026, which we take as the population share of positive sequences, π. Holding recall R and the false-positive rate FPR fixed at test-set values, the implied population precision is $\pi R/[\pi R+(1-\pi)\, FPR]$, from which the F_1 score follows analytically. Columns (5)–(8) of Table 4 report the same metrics as columns (1)–(4), but re-weighted in this way. Again, we find that RWO outperforms all other methods, but in this case the difference in F_1 scores is starker. Our baseline RWO achieves a 0.76 F_1 score, while the F_1 scores of the frontier AI models fall to between 0.65 and 0.70 and other methods are worse. The gap is widest in precision, where RWO attains 0.63 against 0.48–0.55 for the frontier models. Moreover, unsupervised fine-tuning becomes more important as the F_1 score for RWO with generic embeddings drops to 0.68. These results arise because, as Table 4 shows, RWO has a particularly low false positive (FP) rate compared to other methods. When negative examples dominate the evaluation sample, correctly classifying them becomes important for overall performance and RWO is strong in this dimension. Since these re-weighted metrics reflect the label composition of the universe of job postings, our findings highlight the potential gains in accuracy of using our approach.

We view these results as methodologically important because there remain few large-scale exercises for benchmarking the performance of different text classification algorithms in economically relevant tasks. There are two important findings. First, we find notable improvements to using LLMs over simpler, bag-of-words based approaches. Shapiro et al. (2022) do not report large gains from using BERT over simpler models for classifying news sentiment. One reason that we, in contrast, find large gains is the size of our training data. Shapiro et al. (2022) train BERT on 800 labeled articles while our training set is nearly 40 times larger, yielding more information for estimating the complex ways that word sequences map to meaning.

How much of RWO's performance depends on the size of the human-labelled training sample? Figure 1 re-trains RWO on random sub-samples of the labelled training data (five draws at each sample size) and scores every fitted model on the same held-out sample of 4,050 labelled sequences used for Table 4. The error rate falls steeply at first—from 11.2% at 500 labels, no better than the negation-adjusted dictionary's 11.9% on the same sample, to roughly 4% by 3,000 labels—and then flattens, reaching 2.2% by roughly 20,000 labels and improving no further. The returns to additional labelling are therefore largely exhausted well before our full set of 30,000 labels, so a research team facing a similar classification problem could reach most of the attainable performance with a few thousand labels. The plateau sits essentially at the error rate of the best zero-shot frontier model on this sample (GPT-5.5, at

2.0%), although RWO pulls clearly ahead once the metrics are reweighted to the population base rate (Table 4). We expect other text-based prediction problems in economics to benefit similarly from training samples of this scale.

Second, fine-tuning a smallish open-source model like DistilBERT on domain-specific text and human labels matches zero-shot use of much larger generic AI models on this task, at a small fraction of the cost. Scaling zero-shot learning to the full dataset of over three billion sequences would be prohibitively costly. The final column of Table 4 puts the cost at $2.80 per million sequences for RWO against $1,350 for GPT-5.5. At a deeper level, the training and deployment of AI models developed in the private sector are opaque and often not reproducible, especially over long horizons as models are retired from operation. In contrast, our approach is fully open source as early Transformer models like DistilBERT have fully documented training data and downloadable weights. In short, adapting smaller, public models to specific measurement problems is a compelling starting point for empirical economists despite the attention given to large, generic AI models.

A separate question is how RWO-generated statistics compare to those based on other methods. Here, we focus on how RWO compares to the dictionary method, which is most common in the literature measuring remote work in job posting text. Figure A1 plots monthly time series of the share of remote work postings in the US sample from 2019 through early 2023, computed using the same aggregation as in our baseline results in the next section. The two series differ sharply. According to the dictionary method, the remote work share surged at the onset of the COVID-19 pandemic, peaked in early 2021, and fell markedly throughout 2021 before stabilizing in 2022. In contrast, the RWO method suggests a more modest immediate reaction to the pandemic followed by steady growth rate thereafter. Two features of the dictionary series are of note: First, the initial COVID-19 shock drove a large number of both real and negated mentions of remote work arrangements, so this series increases much more dramatically than the RWO series. Second, towards the end of 2022 a handful of very large job boards altered their structure to partially address this issue of negation. Importantly, this second event appears did not induce an abrupt change in our RWO-based measure, suggesting that it is robust to changes in structure that don't alter intended meaning. Clearly, then, the choice of measurement approach can have important quantitative implications even in the aggregate time series.

Of course, aggregate comparisons can mask underlying differences at more granular levels. To illustrate, we compute the change in remote work offerings according to the dictionary method and RWO from 2019 to 2022 for individual SOC2 occupations. As Figure A2 reveals, there are large differences in the specific occupations that each method associates with

growth in remote work adoption over this period. According to the dictionary method, the 'Food Preparation and Serving' occupation experienced the largest change, while for RWO the biggest growth is in 'Computer and Mathematical'. Moreover, according to RWO all occupations experienced an increase in remote-work shares, whereas the share falls for 'Farming, Fishing, and Forestry' occupation according to the dictionary method. The higher accuracy of RWO in the sample of human labels suggests its ranking of occupation-level changes is more reliable.

In sum, the RWO model yields highly accurate classifications and produces markedly different statistics than the most popular alternative based on keyword search. This difference is especially pronounced since 2020, even at the aggregate level.

4 Results

This section proceeds as follows: First, we present information on remote-work shares of new postings over time in our five countries. Second, we compare remote-work shares across occupations, contrasting the outcomes in 2019 and 2025. We also compare our occupation-level outcomes to prominent measures in the literature. Third, we consider city-level variation in remote-work shares in the cross section and over time. Among other things, we relate the rise in remote-work postings to metro-area density and commuting times. Fourth, we compare our measures to survey-based measures derived from the American Community Survey (ACS) and Current Population Survey (CPS). Fifth, we present evidence on differences in remote-work shares across employers operating in the same industry and occupation.

4.1 Remote Work across Countries

How did the share of advertised hybrid and fully remote work differ across countries prior to, during and after the pandemic? In Figure 2 we plot the monthly time series of the share of advertised remote work for the US, UK, Canada, Australia and New Zealand. For each country and in each month, this figure reports the weighted mean of the percent of postings offering remote work across nearly 800 narrow occupation groups. We weight each group based on the share of vacancies in this group in the USA during 2024. Four high-level facts emerge:

1. **Sharp increase of advertised remote work at the onset of COVID-19:** In March-April 2020, the share of new job vacancies which advertised remote work saw a sharp rise across all countries. On average, the share of postings explicitly offering

remote work roughly doubled between February and April 2020. While this immediate increase occurred across all our countries, the level change was most pronounced in countries with a more severe initial COVID outbreak (USA, UK and Canada).

2. **Sustained growth after COVID:** Following the initial spike in early 2020, the share of advertised remote work continued to grow rather than retrace. In level terms, this expansion was most pronounced in the UK—where COVID lockdowns lingered and were relatively severe—rising steadily from roughly 6% in mid-2020 to over 14% by 2023. We also see evidence of delayed acceleration in Australia and New Zealand, where the sharpest growth in remote-work postings occurred later in 2021 as their local pandemic experiences worsened.

3. **Persistence after the pandemic ended:** By June 2026, long after the forcing event of the pandemic and its associated policy mandates ended, the share of remote job postings remains at or near peak levels in the UK, the US and Canada. This persistence is most visible in the UK, which has seen continued growth to reach nearly 17% by the end of our sample. Similarly, shares in the US, Canada, and Australia appear to have stabilized at a high "new normal" of between 9-10%, showing no meaningful sign of reverting to pre-2020 levels. New Zealand shows a more pronounced reversion from its 2023 peak, though still remains well above its pre-pandemic baseline; Australia is also below its peak.

4. **Sizable differences across countries, even before the pandemic:** The USA had the highest advertised remote work share in 2019 at approximately 3.5%. The UK was lower at roughly 2%, whereas Australia, Canada and New Zealand started from lower baselines between 0.8% and 1.5%. By 2026 the spread in levels is much greater—driven largely by the UK's surge—but proportional differences between the US, Canada, and Australia have narrowed as they converge around the 8-10% range.

In our robustness exercises, we also look at the raw shares of remote work, i.e. without the re-weighting applied to our baseline Figure 2. Comparing the unweighted Figure B.2 to Figure 2 tells us the direction and magnitude of the impact that occupation composition plays in our results. As the U.K.–U.S. comparison in Section 2 indicates, the composition adjustment moves levels slightly. It is largest for the UK and Australia and small elsewhere, and the weighted and unweighted monthly series correlate at 0.98 in all five countries.

4.2 Remote Work across Occupations

We first show the share of advertised remote work by grouping job ads into broad occupation groups (based on two-digit SOC 2010 classifications), which Figure 3 reports. For this, we look only at data from the United States. The differences across broad occupation groups vary greatly. In 2019, we see that a small fraction of ads in 'Legal' occupations explicitly offered remote work arrangements, whereas in 2025, this group leads the sample, having grown by 4.7X. Other white-collar professions show similar expansions, for instance 'Computer and Mathematical' occupations grew by 4.6X and 'Business and Financial Operations' by 3.6X. Interestingly, even sectors requiring physical presence show high relative growth from low baselines; 'Food Preparation' postings mentioning remote work multiplied by 27.6X, likely reflecting administrative or hybrid management roles within that sector. As one might expect, the share of advertised remote work correlates positively with computer use, education, and earnings and is lower in occupation groups which require specialised equipment or customer interactions. We further explore this gradient in Appendix B, finding the remote-work share far higher in postings that require a bachelor's degree than a high-school education, and rising with the experience demanded (Figure B.3). Lastly, Figure 3 provides some evidence that the 2019 shares of remote work correlate with post-pandemic shares.

To investigate the relationship between 2019 and 2025 shares further we next turn to an analysis at the detailed ONET occupation-level. We group our US job vacancies into granular occupations (using O*NET definitions), and plot both the 2019 and 2025 percent of advertised remote work (on a log-scale), presented in Figure 4. After dropping a handful of data points with fewer than 250 postings in 2019 or 2025, we retain 716 O*NET occupations.[15] Figure 4 also shows the feasibility classification according to Dingel and Neiman (2020). A black circle represents jobs which these authors classify as 'not suitable for full-time telework', and an orange triangle denotes the opposite.[16] An unweighted ordinary-least-squares trend line is also depicted in blue.

The strong relationship between pre- and post-pandemic remote work adoption is quantified by a bivariate unweighted-OLS fit using a log-log specification, estimated on the 690 of these occupations with non-zero shares in both years. This model yields an R^2 of 0.61, indicating that the share of vacancies advertising remote work in 2019 was strongly predictive of the share in 2025. The estimated slope coefficient suggests an elasticity of 0.76, meaning the

[15] In total, there are 867 O*NET occupations. Our sample of O*NET codes which have greater than 250 vacancy postings in 2019 & 2025 is 716, of which the 690 with non-zero shares in both years enter the log-log fit reported below. This attrition is expected, for example a number of military occupations are not present in our data.

[16] These are taken from the authors replication data, accessed April 2022, which can be found here.

growth has been largely proportional to the initial base.[17] Despite this strong general trend, substantial variation exists. Occupations such as 'Software Developers' and 'HR Managers' sit significantly above the regression line, suggesting adoption rates that outpaced predictions based on historical levels. Furthermore, while the feasibility classification by Dingel and Neiman (2020) accounts for a significant portion of the variation—with 'Teleworkable' jobs (orange triangles) generally clustering in the upper-right quadrant—there are notable discrepancies. For instance, 'Travel Agents' are algorithmically classified as 'not teleworkable' (black circle), yet they appear at the very top of the distribution with high shares of advertised remote work in both periods.[18]

We view three key points of difference between our measurement approach and those measures which assess telework feasibility for each occupation. First, since our measurement works at the job vacancy level and not the occupation level, our measure offers more variation and signals heterogeneity in remote work feasibility within occupations and across firms. Second, whereas the feasibility measures treat each job as a collection of tasks, our measure combines both task-feasibility as well as employer and employee preferences, labour market forces, past experience with remote arrangements, and so on. The third reason for the discrepancy is that our measurement exercise will likely have some amount of under-reporting, as employers may not explicitly advertise remote work in their vacancies but nonetheless allow such arrangements.

Task feasibility only partially predicts where remote work actually appears (Figure B.4). Table B.3 in Appendix B shows that these offers travel with the employer as much as with the task.

4.3 Remote Work across Cities

Next we compare the percent of new vacancy postings which advertised remote work across cities. Job postings are matched to a city based on specific locations listed on the website from which it was scraped, or else mentioned in-text.

[17] Our ordinary least-squares estimates impose a power-law coefficient, given the log-log specification.

[18] In a few cases, the D&N machine classification appears very inaccurate. For example, travel agents have been classified as 'not teleworkable', although both before and after the pandemic roughly 1-in-3 jobs advertised remote work. This is likewise the case for 'Advertising Sales Agents' and 'Interpreters & Translators'. Some of these outliers appear to be resolved by the hand coded measure, but these data are only available at a higher level of occupational-aggregation. The discrepancy also runs the other way: D&N correctly classify "kindergarten teachers" as *feasible for full-time home working (i.e. via a virtual class room), yet teaching jobs in general, and kindergarten teachers in particular, have some of the lowest shares of advertised remote work of any job—anecdotally the arrangement was very taxing on staff and was abandoned as soon as normal schooling resumed, which underlines that feasibility and actual behaviour can vary markedly.*

Our measure pools hybrid and fully remote positions, and the distinction matters for how the geography should be read. Hybrid arrangements keep a worker tethered to the local labour market, since the job still requires attendance at the work site on some days, so the city attached to a posting remains the city in which the work is performed and near which the worker must live. Fully remote positions sever that link and permit long-distance moves, in which case the location attached to the posting identifies the hiring establishment rather than the household of whoever fills the job. The mix of the two margins also differs across occupations. The remote postings we observe in 'Food Preparation' appear to be administrative and supervisory roles attached to a physical site, so their growth reflects locally anchored hybrid arrangements; in 'Computer and Mathematical' occupations a far larger part of the growth plausibly comes from fully remote roles whose holders need not live near their employer at all. Our reading of the labelled postings is that hybrid arrangements are the more common of the two, and we treat the city assigned to a posting as informative on that basis, but this is a reading rather than a measured fact.

Figure 5 shows the percent of advertised remote work across a selection of large international cities, comparing 2019 to 2025. We see that the percentages vary widely. For example, in 2025, roughly 1-in-5 new job postings in Washington (DC) advertised remote work arrangements, compared to roughly 1-in-12 in Adelaide, Australia. The substantial increases as well as the large heterogeneity in these shifts can be seen both across and within countries. Notably, Australian cities have seen some of the largest relative growth in remote-work postings, with Sydney and Melbourne growing 11.5x and 10.4x respectively from their 2019 baselines.

This sizable increase in the levels and spread of remote work across cities, as well as the weak relationship between 2019 and 2025 shares (documented in Figure B.5 of Appendix B), poses an interesting question: What are the city-level determinants of remote work adoption? We hypothesize that a mix of institutional features, infrastructure quality, pandemic severity (both in disease and policy) and the composition of jobs and firms in each city are all important factors. Two city characteristics that proxy for these forces—population density and pre-pandemic commuting time—can be examined directly. Figure 6 sorts the 247 metropolitan areas in our panel for which the American Community Survey publishes 2019 density and commuting characteristics into terciles of population density in Panel A and of mean commuting time in Panel B, both measured in 2019 so that the grouping cannot be an artefact of the shock itself. The terciles were nearly indistinguishable before 2020 and have separated steadily since, with the gap between the densest and least dense thirds widening from 0.3 percentage points to 3.2 and the longest-commute third ending the sample

at 10.6% against 7.6% in the shortest-commute third. Both gradients widened rather than closed after the pandemic, which is what one expects if remote work is taken up hardest where commuting is most costly.

We next turn to more granular monthly time series for selected US cities, shown in Figure 7, where we observe data right up to June 2026. As well as illustrating the granularity of our data, a number of interesting features emerge from these time series. Cities from the North-East and West regions (e.g., San Francisco, Boston, New York) all experienced sharp increases at the outset of the pandemic, but displayed very different growth trajectories subsequently. While San Francisco initially led with peaks exceeding 30% in 2022, Boston has since surged, overtaking other cities to reach approximately 25% by late 2025. We also observe substantial fluctuations over time; for instance, a notable dip occurs across several major cities (including San Francisco, New York, and Denver) around late 2023 to early 2024 before recovering. By contrast, cities in the South show far less growth and lower volatility. Savannah and Miami Beach have remained consistently low, hovering near 5% or below, showing only a modest elevation above their pre-pandemic baselines compared to the dramatic shifts seen in tech hubs. Note that in this exercise, we do not re-weight the data, such that much of the variation across cities is likely to be driven by differences in occupation and industry composition. We leave as future work a mapping from our time-series measures and forcing events, such as shelter-in-place orders.

4.4 RWO Outcomes Compared to the ACS and CPS

Our measurement of remote working utilises new job postings (a flow variable), which is conceptually distinct from the stock of employees working from home. To validate that our vacancy-based measure accurately reflects the economic reality of the labor market, we compare it against two official external benchmarks from the U.S. Census Bureau: the American Community Survey (ACS) and the Current Population Survey (CPS). Figure 8 presents these comparisons across four panels, utilising log-log scales to analyse the relationship between the percent of vacancies offering WFH (our RWO measure) and the share of workers engaged in remote work.

Figure 8 Panels A and B utilise data from the 2024 ACS, focusing on the share of employed persons who report "Worked from home" as their primary commuting method.[19] Panel A compares the WFH share across Metropolitan Statistical Areas (MSAs), revealing a strong relationship with a weighted correlation of 0.697. The regression slope of 1.21 indicates that

[19]Since ACS respondents must select only one box for the method used for the "most of the distance", this measure primarily captures fully remote workers rather than hybrid arrangements.

vacancy postings are highly elastic to local work-from-home rates. When aggregated by occupation in Panel B, the alignment is even stronger, with a weighted correlation of 0.916 and a steeper elasticity coefficient of 1.54.

Figure 8 Panels C and D utilize the 2024 CPS, specifically the share of respondents who engaged in "paid work at home" in the reference week—a measure that captures hybrid work arrangements more effectively than the ACS commute question. Comparing across U.S. States in Panel C, we again find a robust positive relationship with a weighted correlation of 0.855 and a slope of 1.28. Panel D, which analyses the CPS data by occupation, exhibits the highest fidelity in the figure with a weighted correlation of 0.934 and a regression slope of 0.9.

Taken as a whole, the evidence in Figure 8 suggests that our RWO measure of remote working opportunities constructed using the flow of new vacancy postings is a highly robust indicator of the cross-sectional incidence of remote work practices measured from surveys of the stock of employed workers. This fact, along with the high-frequency, highly granular nature of our data has led to a large number of academic and non-academic users of the data product, which is publicly available at WFHmap.com.

The postings data also do not record whether a vacancy is ultimately filled, or how long it remains posted, so advertised terms are best read as offers rather than realised matches. The comparison year here remains 2024, the most recent with complete ACS and CPS microdata.

4.5 Remote Work across Companies

Ultimately, the decision to advertise remote work arrangements is made by each employer who is searching for talent. By and large, workers value the flexibility to work some days remotely, with survey evidence estimating that a typical worker would sacrifice 6% of their salary to receive this amenity (Barrero et al., 2021). Thus, one important reason why employers have increasingly chosen to offer remote work arrangements even after the pandemic is to attract workers. Similarly, remote work arrangements can also lessen the burden of distance and allow firms to recruit for talent in wider geographic areas. Again, this deepens the labour market and may facilitate matching with better candidates. Another reason why we see that employers are offering remote work arrangements in their vacancy listings might be due to learning. Most CEOs comment that mass remote-work of staff would have been unthinkable prior to 2020, yet the forced experimentation during COVID-19 has left many with at least an indifference to such practices and at most tangible evidence of the productivity benefits these bring. Finally, firms—especially those which are expanding quickly—may see remote work arrangements as a way to reduce office space and energy consumption. On

the other hand, the need to adjust internal processes to a fully or partially remote workforce may also inhibit firms from explicitly committing to this work arrangement.

The magnitudes involved are easiest to see in named cases. Figure 9 takes employers in the same industry recruiting in the same occupational category and plots the share of their vacancies that offer remote work in 2019, 2022 and 2025. Panel A of Figure 9 shows the remote-work share of postings by four large aerospace manufacturing firms (NAICS code 3364). We consider only management occupations in this panel. The data reveal a striking divergence in post-pandemic strategies. While both Boeing and Lockheed Martin explicitly offered remote arrangements in roughly half of their postings in 2022, their paths separated sharply by 2025. Lockheed Martin expanded these offers to nearly 100% of management vacancies in 2025. In contrast, Boeing significantly retracted its remote offerings, dropping to roughly 10% in 2025. This substantial reduction could indicate a shift towards return-to-office (RTO) mandates or a deprioritization of remote work flexibility. Northrop Grumman followed a similar pattern of retraction, peaking in 2022 before dropping to under 10% in 2025. Interestingly, SpaceX, which made no explicit offers for such arrangements in 2019 or 2022, began offering remote work in a small fraction (approximately 8%) of listings in 2025.

Turning to the insurance sector, Panel B of Figure 9 shows selected insurance firms that advertise vacancies for workers in the mathematical science occupations. We chose this occupation because it historically has a high national share of remote work vacancy postings. United Health Group displays a consistent upward trajectory, starting with a sizable fraction (approx. 63%) in 2019 and growing to nearly 95% by 2025. Mutual of Omaha exhibits a different pattern; while they mentioned such practices in nearly all vacancies (approx. 97%) in 2022, this share retracted to roughly 79% in 2025. Humana saw steady growth across all three periods, more than quadrupling its 2019 share to reach approximately 80% in 2025.

Finally, Panel C of Figure 9 conducts the same exercise for selected auto manufacturing firms that hire engineers. Almost no explicit offers of remote work were made in 2019. The 2025 data reveals significant volatility compared to 2022. Honda, which led the group in 2022 with roughly 33% of postings offering remote work, drastically reduced this to roughly 6.5% in 2025. This pattern could indicate a similar return-to-office push, distinguishing Honda from its peers who continued to expand flexibility. Conversely, General Motors continued to expand remote opportunities, rising to approximately 36% in 2025, effectively swapping positions with Honda. Ford also saw growth, rising to roughly 16% in 2025. Tesla job postings remain an outlier, making almost no offers of remote work across all observed years.

These named cases are not outliers. Figure 10 plots the distribution of 2025 remote-

work shares across the 31,636 U.S. employers with at least 100 postings, both overall and within the largest two-digit industries. Almost 30% of employers advertise no remote work at all while close to 6% advertise it in more than half of their postings, and a variance decomposition attributes 9.0% of the cross-firm variation to differences between two-digit industries and 91.0% to differences within them. What governs how much remote work an employer advertises is almost entirely something other than its industry.

Figure 11 asks whether those differences persist, assigning each firm to a cohort by the first year in which it advertises remote work after a year of posting without doing so and tracking each cohort's quarterly share in a right-balanced panel that excludes staffing agencies. The cohorts fan out and remain ordered by vintage of adoption in almost every quarter, with the 2019 cohort ending our sample at 14.4% against 2.4% for the 2024 cohort, and the spread remains clearly visible against the all-postings average plotted alongside. The ordering established at adoption survives six years after the onset of the pandemic, which points to a durable organisational characteristic rather than a transitory response to labour market conditions.

Appendix B shows further that our firm-level shares align with the Flex Index, an independent compilation of employers' stated work arrangements built from entirely different inputs than posting text (Table B.5), and that the task-feasibility measures of Dingel and Neiman (2020) only partially predict which employers offer remote work, which travels across a firm's occupations as an employer-level trait (Figure B.4, Table B.3).

5 Conclusion

We develop, test, and apply a language model to track the share of job vacancy postings that offer one or more days of remote work per week in the USA, UK, Canada, Australia and New Zealand. The scale and granularity of our data let us track remote-work shares by industry, occupation, and city in each country at a monthly frequency from 2014 to 2026. We publish our data with regular updates at WFHmap.com.

In classifying the remote-work status of job postings, our language model greatly outperforms dictionaries or a "bag of words" approach. Our language model matches or surpasses the performance of frontier AI models, and it does so at less than one percent of their processing costs. We achieve high-quality measurements by fine tuning a smallish, open-source LLM using 30,000 human-generated classifications. We show that a sizable training sample is essential to achieve performance levels that match those of generic AI models.

We use the statistical output from our language model to uncover several findings. First,

the remote-work share of postings rose steadily from early 2020 through 2023. In contrast, the incidence of remote work among the employed fell sharply after spring 2020, eventually stabilizing at levels well above 2019. Second, almost every occupation saw an increase from 2019 to 2025 in the remote-work share job postings. However, the occupation-level remote-work shares are not well explained by remote-suitability measures based on task and activity descriptions. Third, as of 2025, city-level differences in the remote-work share of postings are much wider than in 2019. Moreover, the 2019 city-level shares have only modest predictive power for the 2025 shares. Metro-area population densities and commuting times help explain which cities experienced the largest gains in the remote-work share of job postings. Fourth, at the employer level, early adopters of WFH show persistently higher remote-work shares of job postings years later.

More broadly, we find great heterogeneity in remote-work shares across countries, industries, occupations, and firms. Indeed, even among firms in the same industry competing for talent in the same occupations, we find large differences in the remote-work shares of jobs on offer. This non-uniformity result suggests that, for many occupations, it is misleading to think of remote-work suitability as a technological constraint. Rather, it reflects choices about job design and how to run an organization. These choices can be influenced by organizational history, management philosophy, workforce characteristics, commuting burdens, government policies, and other features of the external environment. Exploring how these factors shape the evolution of work arrangements is an exciting area for future research.

References

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo (Apr. 2022). "Artificial Intelligence and Jobs: Evidence from Online Vacancies". In: *Journal of Labor Economics* 40 (S1). Publisher: The University of Chicago Press, S293–S340. ISSN: 0734-306X. DOI: 10.1086/718327.

Adams-Prassl, Abi, Maria Balgova, and Matthias Qian (2020). "Flexible Work Arrangements in Low Wage Jobs: Evidence from Job Vacancy Data". In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.3695392.

Adams-Prassl, Abi, Teodora Boneva, Marta Golin, and Christopher Rauh (Jan. 1, 2022). "Work that can be done from home: evidence on variation within and across occupations and industries". In: *Labour Economics* 74, p. 102083. ISSN: 0927-5371. DOI: 10.1016/j.labeco.2021.102083.

Adrjan, Pawel, Gabriele Ciminelli, Alexandre Judes, Michael Koelle, Cyrille Schwellnus, and Tara Sinclair (2021). "Will it stay or will it go? Analysing developments in telework during COVID-19 using online job postings data". In: DOI: https://doi.org/10.1787/aed3816e-en.

Aksoy, Cevat Giray, Jose Maria Barrero, Nicholas Bloom, Steven J. Davis, Mathias Dolls, and Pablo Zarate (June 17, 2022). "Working From Home Around the World". Brookings Papers on Economic Activity.

Alipour, Jean-Victor, Oliver Falck, and Simone Schüller (Jan. 1, 2023). "Germany's capacity to work from home". In: *European Economic Review* 151, p. 104354. ISSN: 0014-2921. DOI: 10.1016/j.euroecorev.2022.104354.

Ash, Elliott and Stephen Hansen (2023). "Text Algorithms in Economics". In: *Annual Review of Economics* 15.1, pp. 659–688. DOI: 10.1146/annurev-economics-082222-074352.

Ash, Elliott, Stephen Hansen, Claudia Marangon, and Yabra Muvdi (2025). "Large Language Models in Economics". In: *Handbook of Economics and Language*. Palgrave Handbooks.

Autor, David, Arindrajit Dube, and Annie McGrew (Mar. 2023). *The Unexpected Compression: Competition at Work in the Low Wage Labor Market*. Working Paper 31010. National Bureau of Economic Research. DOI: 10.3386/w31010.

Bai, John (Jianqiu), Erik Brynjolfsson, Wang Jin, Sebastian Steffen, and Chi Wan (Mar. 2021). *Digital Resilience: How Work-From-Home Feasibility Affects Firm Performance*. DOI: 10.3386/w28588.

Bajari, Patrick, Zhihao Cen, Victor Chernozhukov, Manoj Manukonda, Jin Wang, Ramon Huerta, Junbo Li, Ling Leng, George Monokroussos, Suhas Vijaykunar, and Shan Wan

(Feb. 22, 2021). *Hedonic prices and quality adjusted price indices powered by AI*. Working Paper CWP04/21. Cemmap. DOI: 10.47004/wp.cem.2021.0421.

Bamieh, Omar and Lennart Ziegler (June 1, 2022). "Are remote work options the new standard? Evidence from vacancy postings during the COVID-19 crisis". In: *Labour Economics* 76, p. 102179. ISSN: 0927-5371. DOI: 10.1016/j.labeco.2022.102179.

Bana, Sarah H (2022). *work2vec: Using Language Models to Understand Wage Premia*. Unpublished Manuscript, p. 37.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis (July 20, 2020). *COVID-19 Is Also a Reallocation Shock*. DOI: 10.3386/w27137.

— (Apr. 2021). *Why Working from Home Will Stick*. DOI: 10.3386/w28731.

Barrios, John M, Yael Hochberg, and Hanyi (Livia) Yi (Dec. 2024). *Hustling From Home? Work From Home Flexibility and Entrepreneurial Entry*. Working Paper 33237. National Bureau of Economic Research. DOI: 10.3386/w33237.

Bartik, Alexander W., Zoe B. Cullen, Edward L. Glaeser, Michael Luca, and Christopher T. Stanton (June 2020). *What Jobs are Being Done at Home During the Covid-19 Crisis? Evidence from Firm-Level Surveys*. DOI: 10.3386/w27422.

Bick, Alexander, Adam Blandin, and Karel Mertens (Oct. 2023). "Work from Home before and after the COVID-19 Outbreak". In: *American Economic Journal: Macroeconomics* 15.4, pp. 1–39. ISSN: 1945-7707. DOI: 10.1257/mac.20210061.

Bloom, Nicholas, Gordon B. Dahl, and Dan-Olof Rooth (2026). "Work from Home and Disability Employment". In: *American Economic Review: Insights*. Forthcoming. DOI: 10.1257/aeri.20240538.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying (Feb. 1, 2015). "Does Working from Home Work? Evidence from a Chinese Experiment". In: *The Quarterly Journal of Economics* 130.1, pp. 165–218. ISSN: 0033-5533. DOI: 10.1093/qje/qju032.

Brynjolfsson, Erik, John J. Horton, Adam Ozimek, Daniel Rock, Garima Sharma, and Hong-Yi TuYe (June 2020). *COVID-19 and Remote Work: An Early Look at US Data*. DOI: 10.3386/w27344.

Burke, Mary, Alicia Sasser Modestino, Shahriar Sadighi, Rachel Sederberg, and Bledi Taska (May 28, 2020). *No Longer Qualified? Changes in the Supply and Demand for Skills within Occupations*. Federal Reserve Bank of Boston Research Department Working Papers. Series: Federal Reserve Bank of Boston Research Department Working Papers. Federal Reserve Bank of Boston. DOI: 10.29412/res.wp.2020.03.

Council of Economic Advisers (Mar. 2024). *Economic Report of the President, 2024*.

— (Jan. 2025). *Economic Report of the President, 2025*.

Cowan, Benjamin W and Kairon Shayne D Garcia (Aug. 2024). *Remote Work and Political Preferences: Evidence from U.S. Counties*. Working Paper 32834. National Bureau of Economic Research. DOI: 10.3386/w32834.

Criscuolo, Chiara, Peter Gal, Timo Leidecker, Francesco Losma, and Giuseppe Nicoletti (2021). "The role of telework for productivity during and post-COVID-19". In: 31. DOI: https://doi.org/https://doi.org/10.1787/7fe47de2-en.

Davis, Steven J., Cevat Giray Aksoy, Jose Maria Barrero, Nicholas Bloom, Katelyn Cranney, Mathias Dolls, and Pablo Zarate (Mar. 2026). *Work from Home and Fertility*. Working Paper 34963. National Bureau of Economic Research.

Davis, Steven J. and Brenda Samaniego de la Parra (July 10, 2020). *Application Flows*.

Deming, David and Lisa B. Kahn (Jan. 2, 2018). "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals". In: *Journal of Labor Economics* 36 (S1). Publisher: The University of Chicago Press, S337–S369. ISSN: 0734-306X. DOI: 10.1086/694106.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (June 2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. NAACL-HLT 2019. Minneapolis, Minnesota: Association for Computational Linguistics, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

Dingel, Jonathan I. and Brent Neiman (Sept. 1, 2020). "How many jobs can be done at home?" In: *Journal of Public Economics* 189, p. 104235. ISSN: 0047-2727. DOI: 10.1016/j.jpubeco.2020.104235.

Draca, Mirko, Emma Duchini, Roland Rathelot, Arthur Turrell, and Giulia Vattuone (June 16, 2022). *Revolution in Progress? The Rise of Remote Work in the UK*.

Forsythe, Eliza, Lisa B. Kahn, Fabian Lange, and David Wiczer (Sept. 1, 2020). "Labor demand in the time of COVID-19: Evidence from vacancy postings and UI claims". In: *Journal of Public Economics* 189, p. 104238. ISSN: 0047-2727. DOI: 10.1016/j.jpubeco.2020.104238.

Hershbein, Brad and Lisa B. Kahn (July 2018). "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings". In: *American Economic Review* 108.7, pp. 1737–1772. ISSN: 0002-8282. DOI: 10.1257/aer.20161570.

Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean (Mar. 2015). *Distilling the Knowledge in a Neural Network*. DOI: 10.48550/arXiv.1503.02531. arXiv: 1503.02531 [cs, stat].

Huang, Allen H., Hui Wang, and Yi Yang (2023). "FinBERT: A Large Language Model for Extracting Information from Financial Text". In: *Contemporary Accounting Research* 40.2, pp. 806–841. DOI: 10.1111/1911-3846.12832.

Lambert, Peter John and Yannick Schindler (May 18, 2026). *The Broken Ladder: AI, Remote Work, and Early-Career Hiring*. SSRN working paper 6787638. DOI: 10.2139/ssrn.6787638.

Marinescu, Ioana and Ronald Wolthoff (Apr. 2020). "Opening the Black Box of the Matching Function: The Power of Words". In: *Journal of Labor Economics* 38.2. Publisher: The University of Chicago Press, pp. 535–568. ISSN: 0734-306X. DOI: 10.1086/705903.

Mas, Alexandre and Amanda Pallais (Dec. 2017). "Valuing Alternative Work Arrangements". In: *American Economic Review* 107.12, pp. 3722–3759. ISSN: 0002-8282. DOI: 10.1257/aer.20161500.

Modestino, Alicia Sasser, Daniel Shoag, and Joshua Ballance (Aug. 1, 2016). "Downskilling: changes in employer skill requirements over the business cycle". In: *Labour Economics*. SOLE/EALE conference issue 2015 41, pp. 333–347. ISSN: 0927-5371. DOI: 10.1016/j.labeco.2016.05.010.

Mongey, Simon, Laura Pilossoph, and Alexander Weinberg (Sept. 1, 2021). "Which workers bear the burden of social distancing?" In: *The Journal of Economic Inequality* 19.3, pp. 509–526. ISSN: 1573-8701. DOI: 10.1007/s10888-021-09487-6.

Ozimek, Adam (May 27, 2020). *The Future of Remote Work*. Rochester, NY. DOI: 10.2139/ssrn.3638597.

Pfeifer, Moritz and Vincent P. Marohl (2023). "CentralBankRoBERTa: A fine-tuned large language model for central bank communications". In: *The Journal of Finance and Data Science* 9, p. 100114. DOI: 10.1016/j.jfds.2023.100114.

Phuong, Mary and Marcus Hutter (July 19, 2022). *Formal Algorithms for Transformers*. DOI: 10.48550/arXiv.2207.09238. arXiv: 2207.09238[cs].

Rio-Chanona, R Maria del, Penny Mealy, Anton Pichler, François Lafond, and J Doyne Farmer (Sept. 28, 2020). "Supply and demand shocks in the COVID-19 pandemic: an industry and occupation perspective". In: *Oxford Review of Economic Policy* 36 (Supplement_1), S94–S137. ISSN: 0266-903X. DOI: 10.1093/oxrep/graa033.

Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf (Feb. 29, 2020). "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". In: *arXiv:1910.01108 [cs]*. arXiv: 1910.01108.

Shapiro, Adam Hale, Moritz Sudhof, and Daniel J. Wilson (June 1, 2022). "Measuring news sentiment". In: *Journal of Econometrics* 228.2, pp. 221–243. ISSN: 0304-4076. DOI: 10.1016/j.jeconom.2020.07.053.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin (2017). "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc.

Wiswall, Matthew and Basit Zafar (Feb. 1, 2018). "Preference for the Workplace, Investment in Human Capital, and Gender*". In: *The Quarterly Journal of Economics* 133.1, pp. 457–507. ISSN: 0033-5533. DOI: 10.1093/qje/qjx035.

TABLES AND FIGURES

Table 1: Counts of Vacancy Postings, Employers, and Cities, January 2014 to June 2026

	(1)	(2)	(3)	(4)
Country	Vacancies	Employers	Cities	Sources
United States	437,928,326	8,747,064	47,419	57,088
United Kingdom	113,043,254	2,372,824	18,078	21,805
Canada	28,303,400	1,580,084	11,529	14,748
Australia	15,571,361	556,426	1,231	8,312
New Zealand	3,928,789	158,243	218	4,282
All	598,775,130	13,414,641	78,475	106,235

Note: Reported counts pertain to the universe of online postings from January 2014 to June 2026, inclusive. We rely on our vacancy data providers proprietary algorithm to identify employers and cities. Column (4) reports the number of unique web domains from which postings were collected, which include private online job vacancy aggregators, public online job boards, and employers' own recruitment web pages. 'All' row reports column sums.

Table 2: Examples of Classification Errors in Dictionary Methods

False-Positive Examples:	False-Negative Examples:
We are looking for a Deputy Home Manager with domiciliary care experience to join our company. You will **work from home** care facilities with a strong track record of quality service.	We encourage our people to explore new ways of working - including part-time, job-share or working from different kinds of locations, including their home. Everyone can ask about it.
Schedule: * 10 Hour Shift * 8 Hour Shift **Work remotely**: * No	With a hybrid mix of time at home as well as our corporate office, this role will suit an analytical, process orientated and people focused payroll professional who thrives in a fast-paced environment.
Applicants must also have: * Ability to work as part of a team, in a fast paced environment * Experience in a 4 or 5 star hotel * Previous experience working in **remote** locations	We see the value in work-life balance, so whether you like to get a surf in before work, like to head home in time to pick up the kids or you just like working from the comfort of your own home now and then, we want to support you.
You may work on renovation projects, store reorganizations, new store openings, and store closings. May respond to managerial or **Home Office** requests for special reports, information, or for help on special projects.	The interviews for this role are likely to be conducted remotely using Microsoft Teams or Zoom. It is also expected that relevant work within these roles may be done remotely, within the UK.

Note: The left column provides examples of how a dictionary method falsely classifies a vacancy posting as saying the job allows remote work. The right column shows examples of how it falsely classifies a vacancy posting as not saying that the job allows remote work. Bold font designates dictionary keywords, and yellow shading highlights text that helps determine a correct classification. These examples are based on actual vacancy postings in our dataset and the dictionary used in Adrjan et al. (2021).

Table 3: Attention Weights from the RWO large language model, as compared to the Dictionary Keywords

RWO View:	Dictionary View:
Schedule: ** 10 Hour Shift ** 8 Hour Shift Work remotely: ** No	Schedule: ** 10 Hour Shift ** 8 Hour Shift Work remotely: ** No
We are looking for a Deputy Home Manager with domiciliary care experience to join our company. You will work from home care facilities with a strong track record of quality service.	We are looking for a Deputy Home Manager with domiciliary care experience to join our company. You will work from home care facilities with a strong track record of quality service.
We encourage our people to explore new ways of working - including part-time, job-share or working from different kinds of locations, including their home. Everyone can ask about it.	We encourage our people to explore new ways of working - including part-time, job-share or working from different kinds of locations, including their home. Everyone can ask about it.

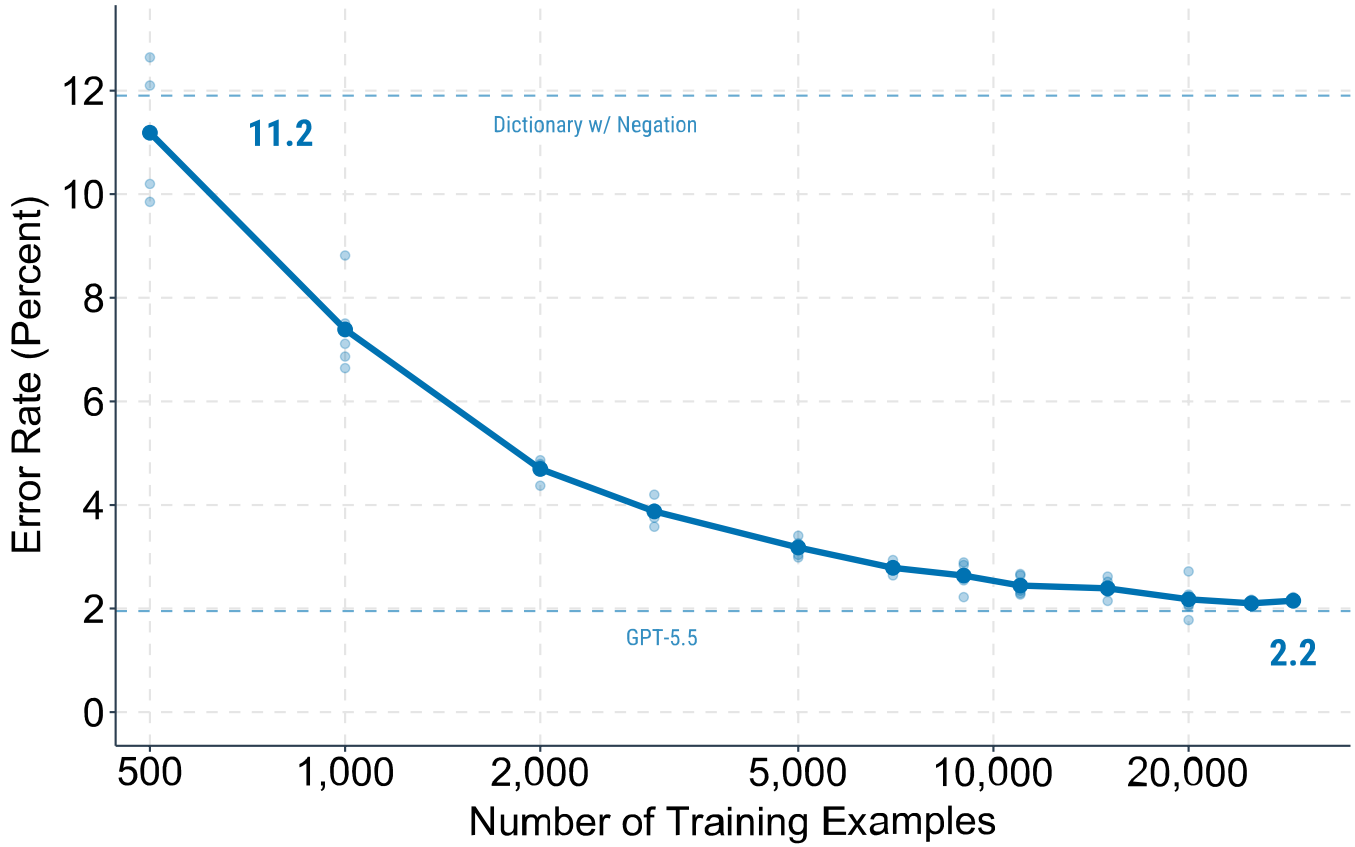
Note: The left column illustrates the role of attention weights in our 'Remote Work in job Openings' (RWO) large language model classifications of vacancy postings, where darker shadings pertain to higher weights. The right column illustrates the application of dictionary methods to the same text passages, where highlight text pertains to keywords.

Table 4: RWO Performance Comparable to Frontier AI Model Classification Methods, at Much Lower Cost

	Audit Sample				Reweighted Sample				Cost
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Error Rate	Recall	Precision	F1 Score	Error Rate	Recall	Precision	F1 Score	USD$/1M
RWO (Baseline)	0.02	0.97	0.97	0.97	0.01	0.97	0.63	0.76	$2.80
GPT-5.5	0.02	0.98	0.95	0.97	0.02	0.98	0.55	0.70	$1,350
Claude Opus 4.8	0.03	0.93	0.95	0.94	0.02	0.93	0.55	0.69	$994
RWO (no job ad pre-training)	0.03	0.96	0.95	0.95	0.02	0.96	0.52	0.68	-
GPT-4o	0.03	0.94	0.95	0.94	0.02	0.94	0.52	0.67	$283
Claude Sonnet 4.6	0.02	0.97	0.94	0.96	0.02	0.97	0.48	0.65	$436
Logistic Regression w/ Negation	0.08	0.83	0.87	0.85	0.05	0.83	0.28	0.42	-
Logistic Regression	0.11	0.81	0.81	0.81	0.07	0.81	0.21	0.33	-
Dictionary w/ Negation	0.12	0.74	0.82	0.78	0.07	0.74	0.21	0.33	-
Dictionary	0.16	0.81	0.68	0.74	0.15	0.81	0.11	0.20	-
All Zero	0.28	0.00	0.00	0.00	0.02	0.00	0.00	0.00	-

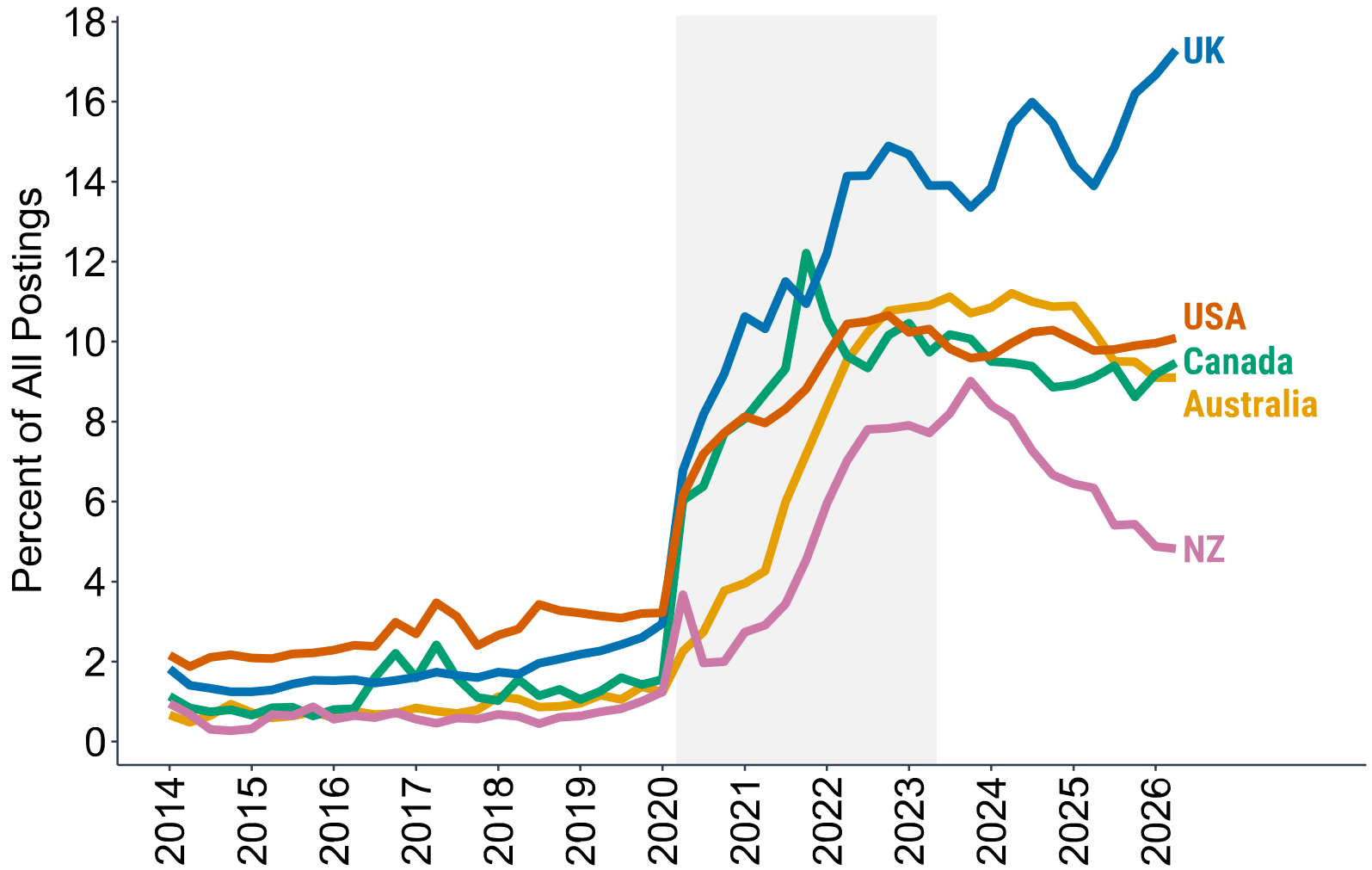
Note: This table reports classification performance metrics, which we calculate using a hold-out sample of human-classified text sequences. “Error Rate” is the overall rate of misclassifications (relative to humans). “Precision” is the ratio of true-positive classifications to the sum of true positives and false positives. “F1 score” is the harmonic mean of Precision and “Recall”, where Recall is the fraction of true positives divided by the sum of true positives and false negatives – i.e., the denominator is the true number of positives, according to human classifications. Columns (1)-(4) are computed on our audit sample; columns (5)-(8) reweight these metrics to the positive-class share of text sequences in our full universe of postings over January 2022 to June 2026 (2.27%). Cost is calculated based on either API price (for frontier models) or else the variable cost of GPU resources (based on GPU, CPU, Memory, and Disk costs using a Google Cloud server, priced in 8th July, 2026). See Appendix C for details, including a description of each algorithm.

Figure 1: RWO Classification Performance Rises Steeply with Training-Sample Size, then Flattens



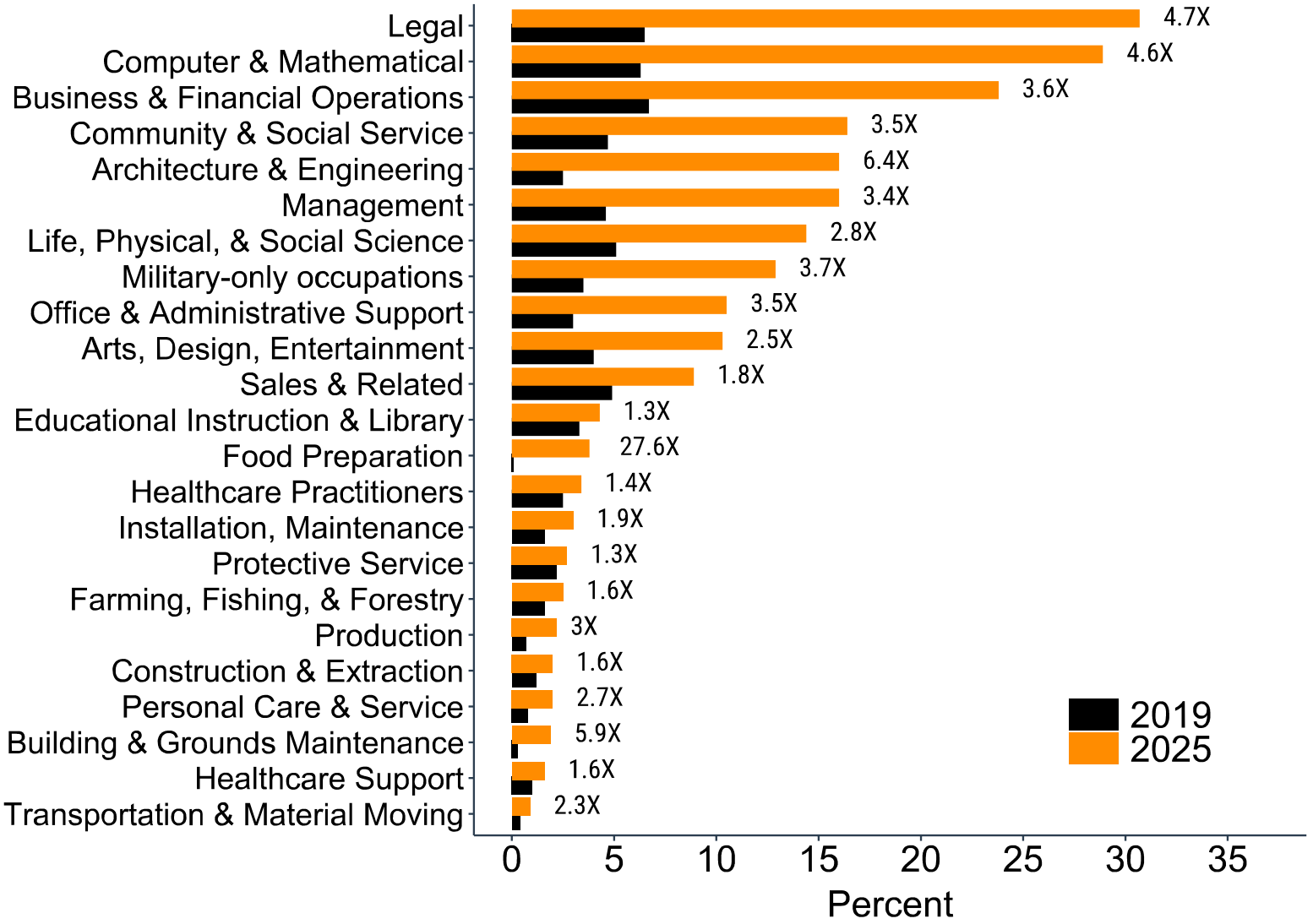
Note: Each point re-trains RWO on a randomly drawn sub-sample of the human-labelled training data (x-axis, log scale; five draws per size) and scores it on the same held-out sample of 4,050 labelled sequences used for Table 4. Points show the individual draws, the line their mean. Dashed lines mark the error rates of Dictionary w/ Negation and GPT-5.5 on the same sample (Table 4). At a training sample of 500 labels - the scale of the training data in Shapiro et al. (2022) - the fine-tuned model's error rate (11.2%) is no better than the dictionary baseline (11.9%); returns are largely exhausted beyond roughly 20,000 labels.

Figure 2: Vacancy Postings that Explicitly Offer Hybrid or Fully Remote Work Rose Sharply and Persistently



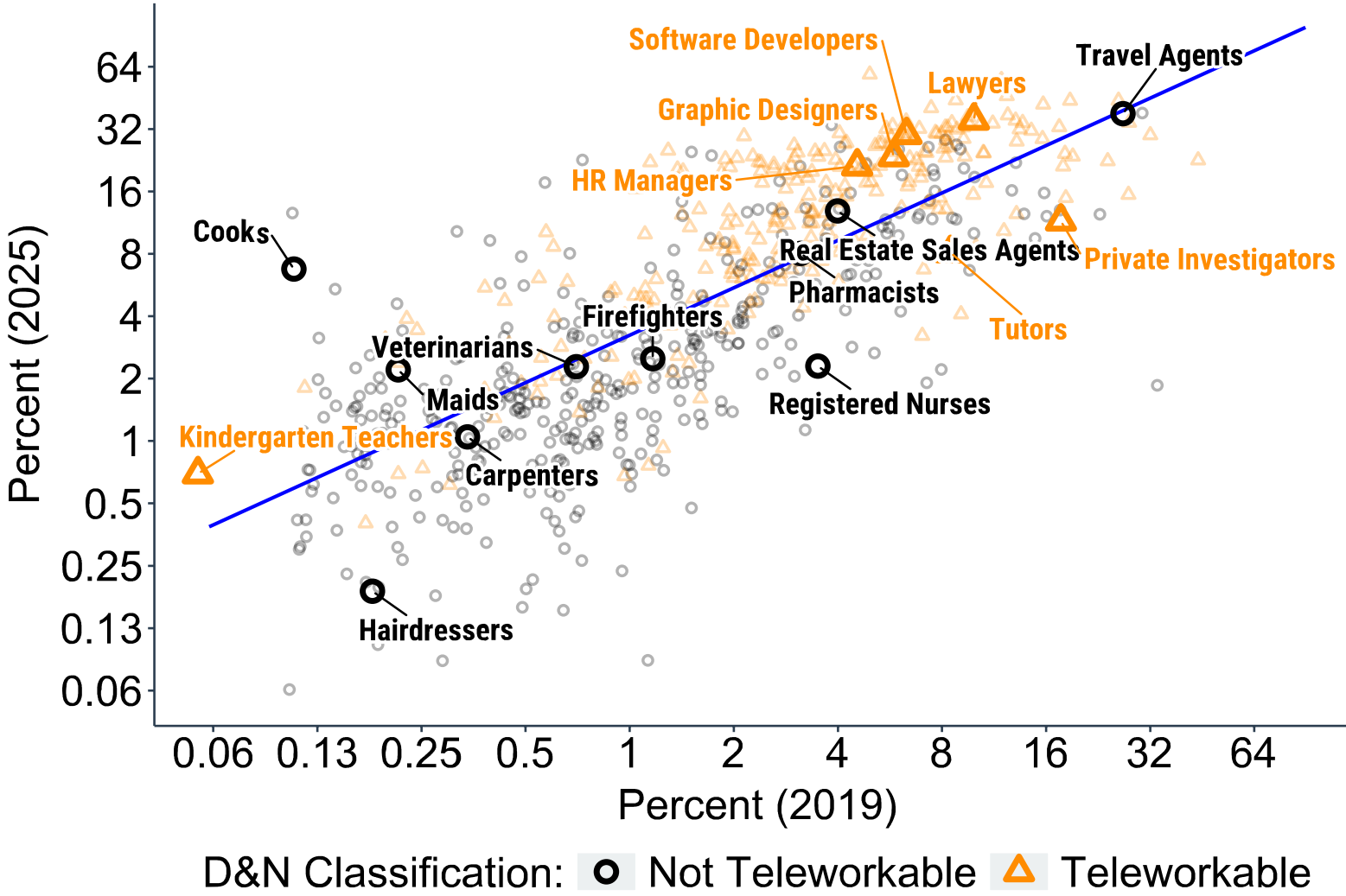
Note: This figure shows the percent of vacancy postings that say the job allows one or more remote workdays per week, encompassing both hybrid and fully-remote working arrangements. We compute these monthly, country-level shares as the weighted mean of the own-country occupation-level shares, with weights given by the U.S. vacancy distribution in 2024. Our occupation-level granularity is roughly equivalent to four-digit SOC codes. See Appendix B for the corresponding raw series and series based on alternative weighting schemes.

Figure 3: Professional, Scientific and Computer-Related Occupations Have the Highest Shares of Postings that Offer Hybrid or Fully-Remote Work, U.S. Data



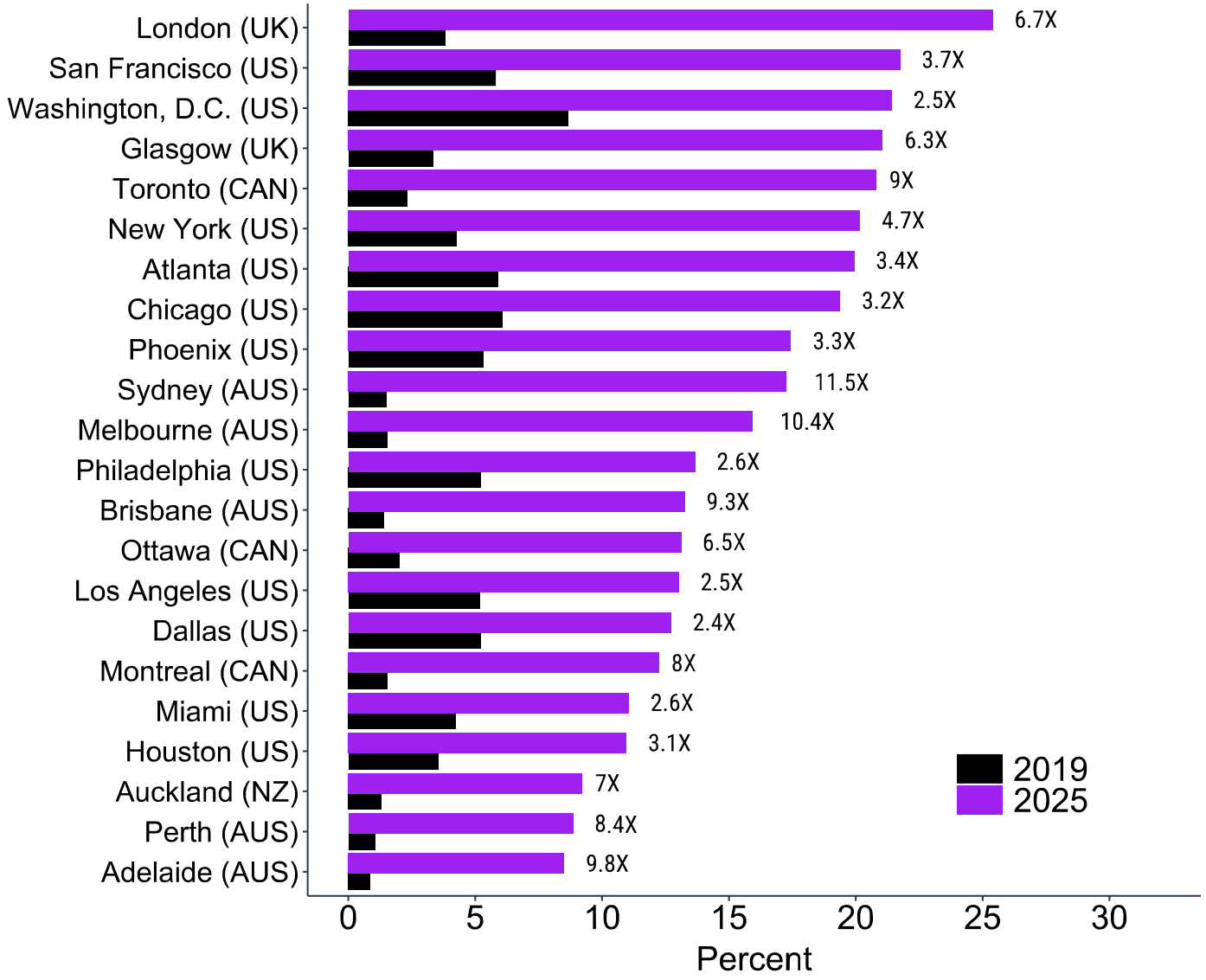
Note: Each bar reports the percent of vacancy postings that say the job allows one or more remote workdays per week in the indicated period and occupation group (two-digit SOC).

Figure 4: The Share of Vacancy Postings that Explicitly Offer Hybrid or Fully Remote Work Rose in Almost Every Occupation, U.S. Data



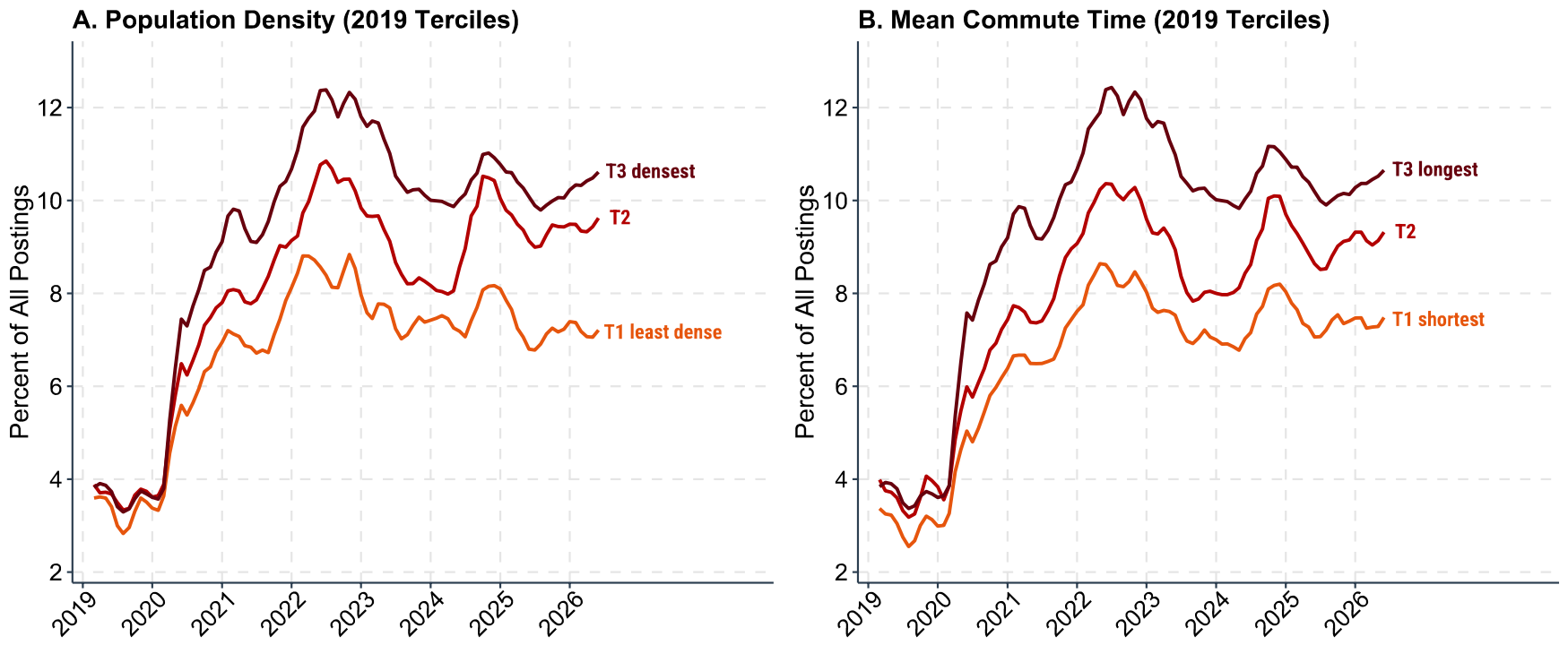
Note: This figure plots the percent of postings that say the job allows one or more remote workdays per week for 690 occupations in 2019 and 2025. We define occupations by ONET codes, dropping those with fewer than 250 posting in 2019 or 2025. The line shows the unweighted OLS fit: log(y) = 1.18 + 0.76 log(x), which has an R^2 value of 0.61. The color and shape denote whether Dingel & Neiman (2020) classify the occupation as feasible for fully remote working.

Figure 5: The Share of Vacancy Postings that Explicitly Offer Hybrid or Fully Remote Work Varies Widely across Major Cities



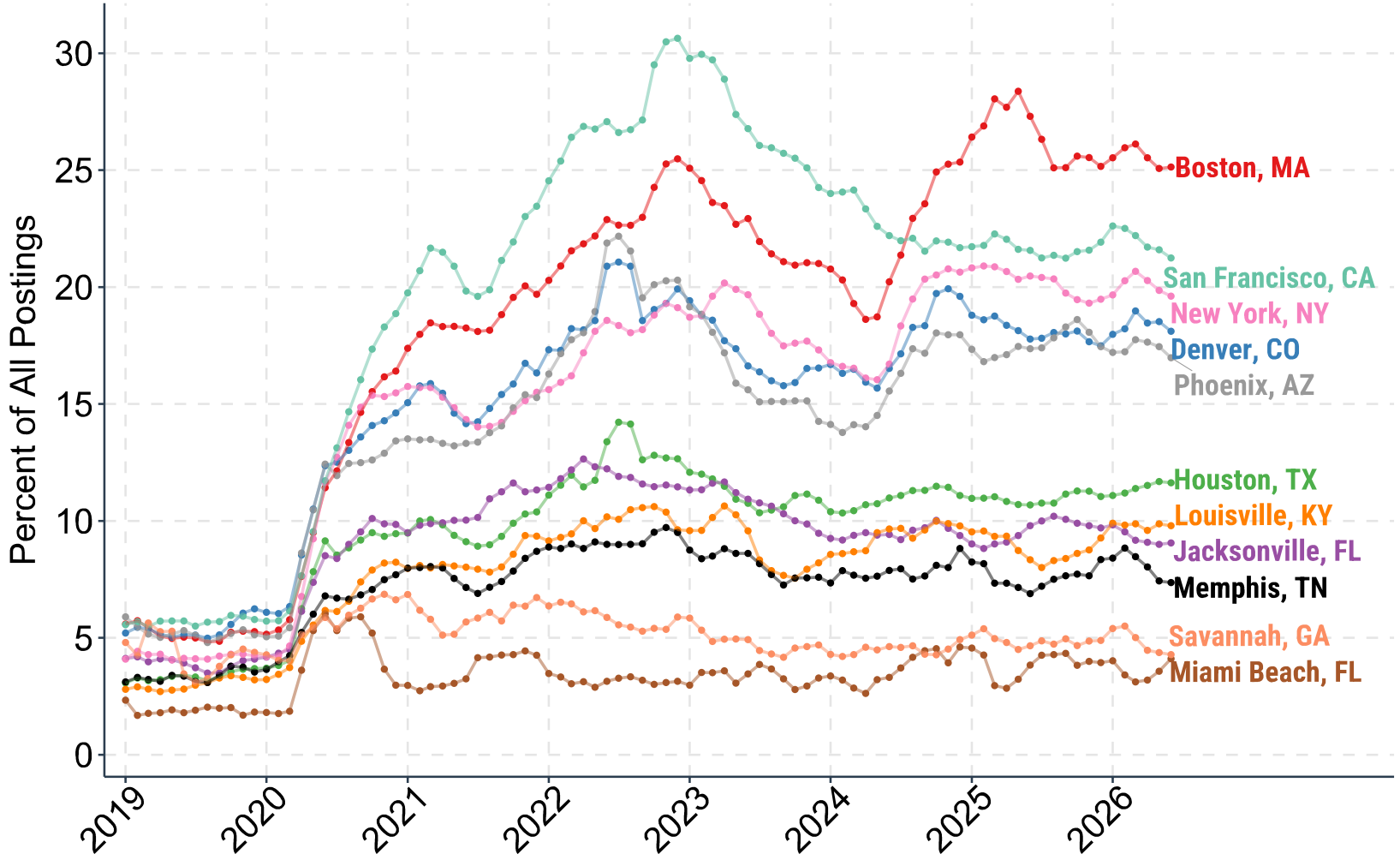
Note: Each bar reports the percent of vacancy postings that say the job allows one or more remote workdays per week in the indicated period and city. City refers to the location of the establishment or firm that is hiring.

Figure 6: Remote Work Rose Most, and Stayed Highest, in Dense, Long-Commute Cities



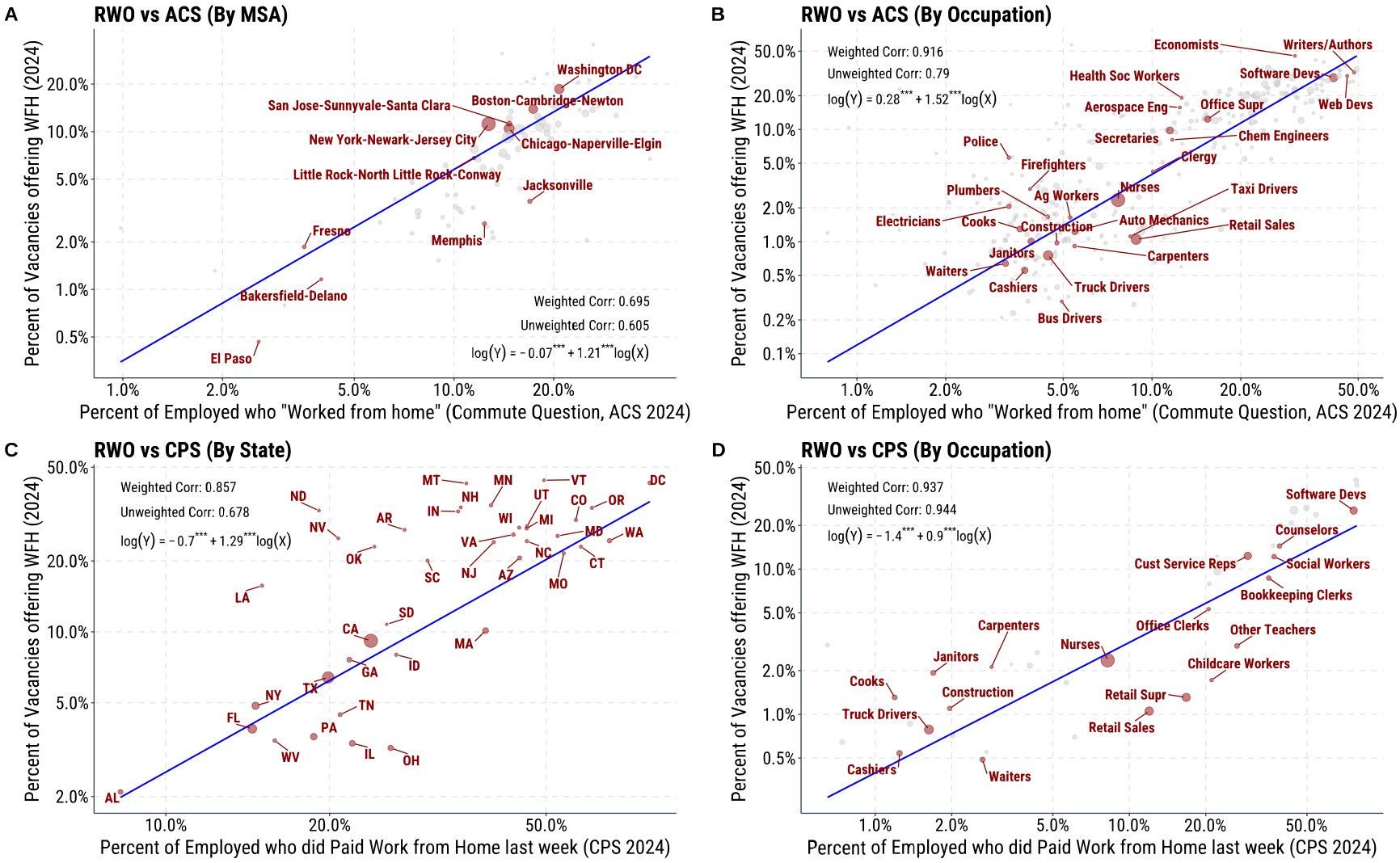
Note: Share of U.S. postings that offer one or more remote workdays per week, by groups of metropolitan areas (CBSAs), January 2019 to June 2026. Cities are sorted into terciles on their PRE-PANDEMIC (2019 ACS) characteristics - population density (Panel A) and mean commute time (Panel B) - so the grouping cannot be an artefact of the remote-work shock itself. Both panels plot monthly posting-weighted group shares smoothed with a three-month trailing moving average. The terciles were nearly indistinguishable before 2020: the densest-third minus least-dense-third gap widened from 0.3 to 3.2 percentage points by June 2026, and the longest- minus shortest-commute gap from 0.6 to 3.0.

Figure 7: Share of Postings Offering Hybrid or Fully Remote Work vary across US cities



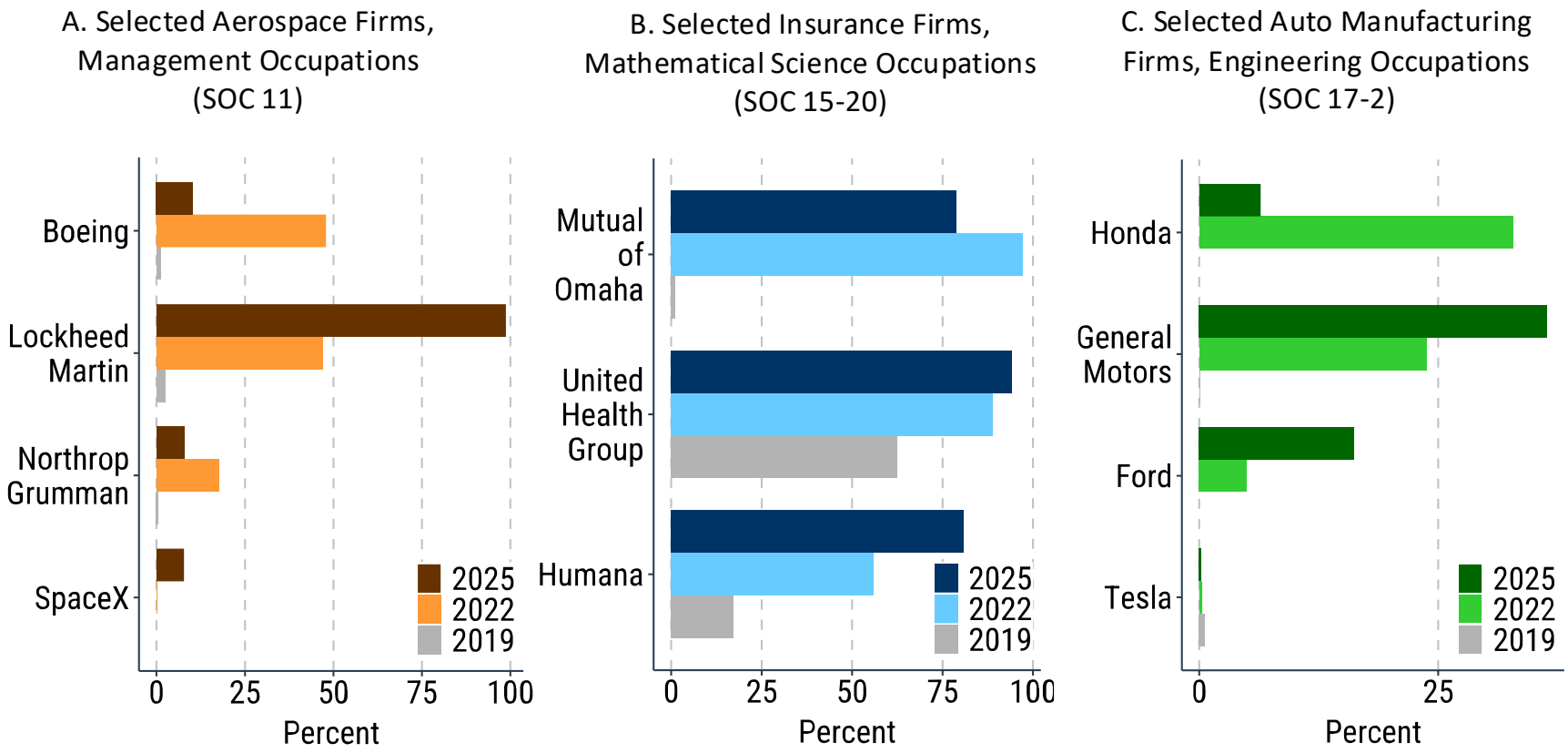
Note: We calculate the monthly share of all new job vacancy postings which explicitly advertise remote working arrangements (i.e. both hybrid and fully-remote), by selected cities. This figure shows the 3-month moving average. Cities chosen above are selected examples to illustrate the wide cross-city spread.

Figure 8: Our 'Remote Work in job Openings' (RWO) measure strongly correlates with Census Bureau surveys (ACS & CPS) across both Occupation and Geography, 2024



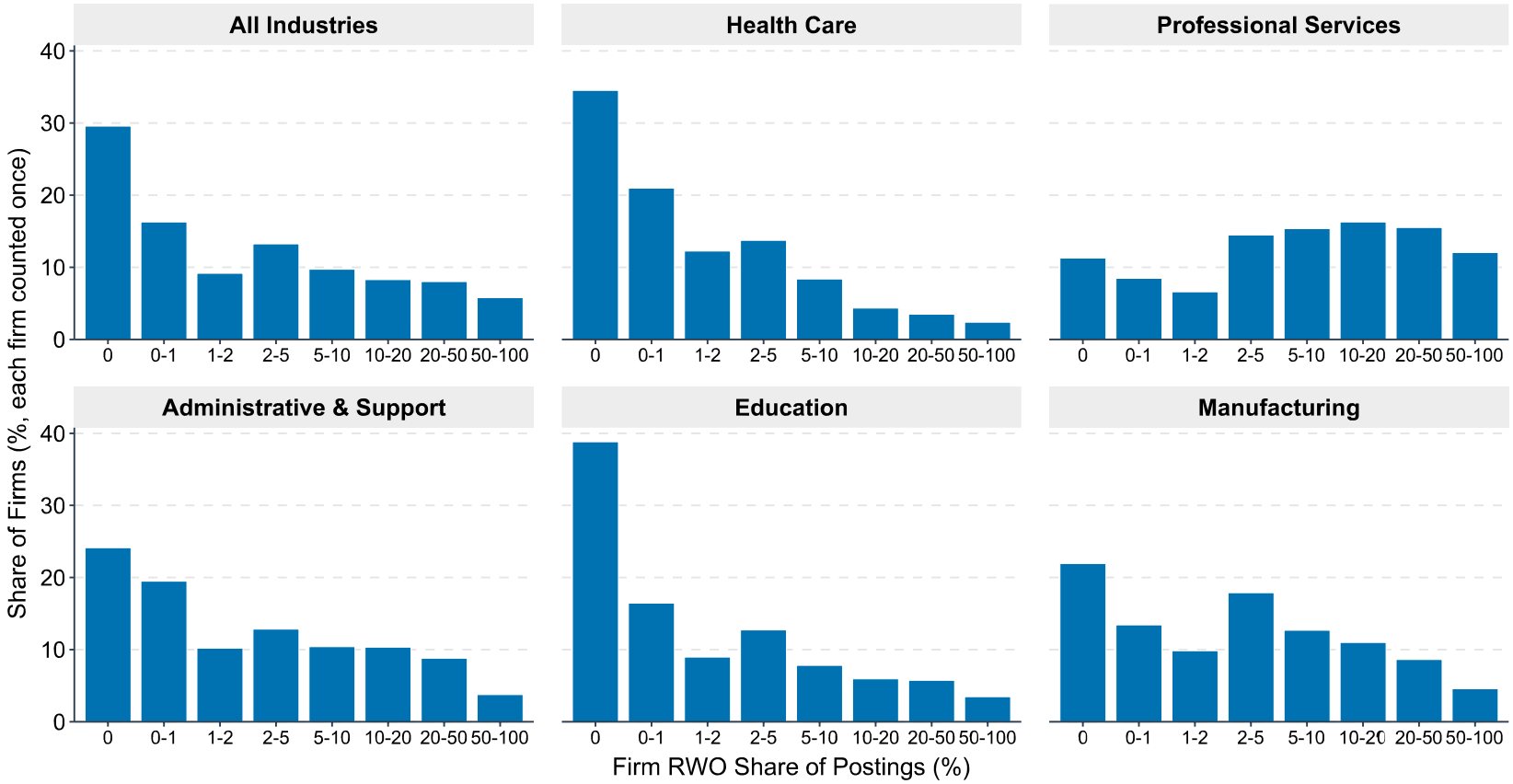
Note: Panels A and B compare our 'Remote Work in job Openings' (RWO) large language model measurements to the share of ACS 2024 respondents who say their primary commute mode is "work from home". Panels C and D compares RWO to CPS 2024 respondents who say they engaged in paid work at home in the last week. All axes use log scales. Blue lines represent unweighted OLS fits.

Figure 9: The Prevalence of Postings that Allow Hybrid or Fully-Remote Work Varies Greatly, even among Same-Industry Firms Recruiting in the Same Occupational Category



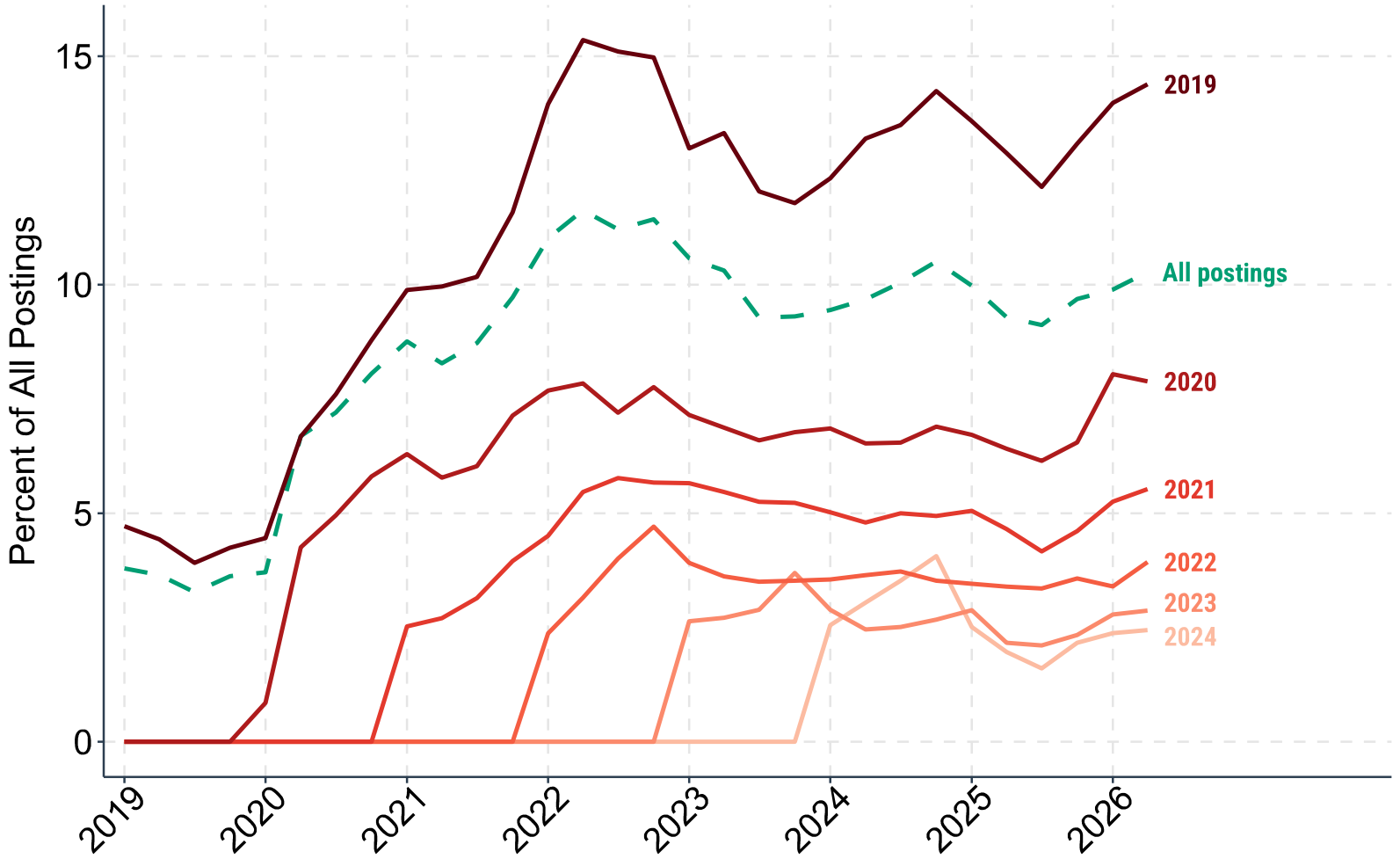
Note: For each firm, year and indicated occupation, we report the percent of U.S. postings that say the job allows one or more remote workdays per week.

Figure 10: Remote Work Differs Enormously across Firms, Even within the Same Industry



Note: U.S. postings, 2025. The sample is the 31,636 employers with at least 100 postings and a measurable remote-work share; a firm's RWO share is the percent of its postings that say the job allows one or more remote workdays per week. The distribution is across FIRMS: each firm counts once whatever its posting volume, so the bars describe the typical employer rather than the typical job, and the variance decomposition below is likewise unweighted. The first panel shows the distribution across all firms; the remaining panels repeat it within the five largest sectors (two-digit NAICS). Bins widen with the share (0, 0-1, 1-2, 2-5, 5-10, 10-20, 20-50, 50-100 percent) because the distribution is highly skewed: 30% of firms advertise no remote work at all, while a long tail advertise it in most postings. Industry explains little of the variation: 9.0% of the cross-firm variance in RWO shares lies between two-digit industries and 91.0% within them.

Figure 11: Early Adopters of WFH Have Persistently Higher Shares of Postings Offering Remote Work Arrangements, compared to Companies which Adopt Later



Note: Figure splits companies into cohorts based on the first year we observe a company offer remote work arrangements in at least one job posting ("adoption"), following a year where the company posted job openings without WFH. For each cohort, we report the unweighted share of WFH job postings pooled across all companies in each quarter. Sample is right-balanced: we exclude firms who don't post at least 5 vacancies a year subsequent to entry, but may enter in any year. The total number of companies (postings) in each cohort are 2019: 20,658 (134.1M), 2020: 21,430 (25.5M), 2021: 13,272 (11.7M), 2022: 7,458 (5.2M), 2023: 4,914 (2.5M), 2024: 3,667 (1.4M). The attrition is by construction. We exclude staffing/agency employers. The green dashed line shows the share of remote-work postings across all U.S. postings.

ONLINE APPENDIX

A Data Appendix

In this Appendix we provide further commentary on the corpus of online job vacancy postings.

A.1 Data Provider

Our corpus of online job vacancy postings is provided by the labour market and analytics company 'Lightcast' (formerly Emsi Burning Glass). Lightcast has been scraping online job vacancy postings in the USA since 2007, and has continued to expand to other countries.

A.2 Web Sources

Each job vacancy posting is scraped by Lightcast from the internet. Specifically, the company scrapes over 50,000 web sources. These sources include private online job vacancy aggregators (e.g. Indeed.com, Monster.com), public online job boards (e.g. New York City Department of Labour's 'JobZone'), and employers' own recruitment web pages (e.g. careers.microsoft.com, usajobs.gov). Lightcast actively audits their list of web sources to ensure data from new websites is on-boarded in a timely manner.[20] One of the main competitive advantages of Lightcast's data product is the breadth of their sources. These data are often referred to in the literature as the 'near universe' of online job vacancy postings.

A.3 What's in the job vacancy posting data?

Once an online job vacancy posting is scraped, Lightcast processes this data to produce three categories of information: (i) plain text, (ii) meta data, and (iii) structured data. A description of each of these categories follows presently:

A.3.1 Plain Text

The plain text of each job ad contains the full textual description of the job, as written by employers. To construct this, Lightcast takes the HTML file scraped from a given website and does two further processing steps. First, it parses out portions of the HTML file which do not contain information about the vacancy (e.g. removing website headers, footers, and side-menu bars). Second, Lightcast takes this portion of HTML which (ideally) contains only information about the job vacancy, and turns this HTML into plain-text.

A.3.2 Meta Data

Each vacancy posting also contains a number of meta-data items. These are immutable properties of each web scraped vacancy. The most important of these is the date the page was scraped. Another important piece of meta-data is the URL from which the posting was scraped.

[20] One reason we eschew analysis of the count of postings and instead focus on shares is that the underlying donor pool of online sources is constantly changing.

A.3.3 Structured Data

The most commonly used data product that Lightcast creates is the set of structured data. This dataset contains one row for each job vacancy posting, and a large number of additional information such as the job title, occupation, salary, educational requirements, location, and employer name. These variables are extracted using Lightcast's own proprietary algorithms. These fields differ from meta data because they may contain missing values and/or measurement error due to imperfect algorithmic extraction.

A.4 Errors and Missing Information

Overall, the data product is a highly informative and accurate product. We view the incidence of errors as very minute, but acknowledge that any dataset with hundreds of millions of observations scraped from over 50,000 sources will never be perfect. Both the structured data and the plain text data require a number of pre-processing steps and the use of algorithmic feature extraction, which in a very small number of cases produce errors (e.g. misclassification of occupations, truncation of plain text, presence of erroneous text). In this subsection we highlight some of the errors we have encountered, and discuss the strategies we employed to ensure our results remain robust to such issues.

A.4.1 Missing Values

A specific value (e.g. the educational requirement for a job) might be missing for at least two reasons: (i) the employer does not mention this explicitly in the text of the job ad, and (ii) the algorithm used to extract this feature from the text failed. The former issue is especially problematic in the context of educational requirements (e.g. we see that very few vacancies for Medical Doctors explicitly mention a requirement to have gone to medical school). This is because certain features of the job will likely be taken as given (for example, specialized degrees for medical doctors). We also see that a large share of vacancy postings do not list the salary (this is almost entirely due to lack of information, and not poor feature extraction). One could employ imputation methods to address these missing values (see Bana (2022), who predicts the salary with a very high degree of accuracy from the text). The main strategy employed in this paper was to only utilise covariates which contain fewer missing values, such as occupation classifications and location information.

A.5 Representativeness of Online Job Vacancy Postings

Lightcast frequently reviews the representativeness of the job vacancy postings it scrapes, to ensure the information renders an accurate picture of the overall labour market. Both our analysis and that of our data provider, as well as many other papers in the literature that utilise these data, all find a high degree of fidelity between the share of job vacancies across occupations and industries, and other official Government data products which measure similar phenomena.

In our baseline results, we also re-weight the data to reduce sensitivity to shifts in the overall composition of the labour market; Section 2.1 sets out the motivation, and Table B.4 reports the alternative schemes.

B Supplementary Results

This appendix gathers additional exhibits on our measurement of remote work. Figure B.1 plots RWO's predicted probabilities over every sequence in the five-country corpus, with 96.3% of sequences in the lowest bin and only 1.2% scoring between 0.1 and 0.9, so the 0.5 threshold cuts through a nearly empty region. Table B.1 reports the number of sequences per job posting which receive a positive classification using 2021 information, where 91.0% contain no positive sequence and a further 5.7% contain exactly one. Table B.2 records the hardware, run time and cost of the two training stages of RWO. Figure B.2 reproduces Figure 2 in the raw posting series, without the occupation reweighting employed in our headline cross-country timeseries.

Job-level results

Job postings typically list a host of requirements for potential applicants. Figure B.3 groups U.S. postings by the minimum education required in Panel A and the minimum experience required in Panel B, comparing 2019 with 2024. Both gradients are steep in the low-to-middle requirement groups. In 2024, 5.2% of postings asking for no more than a high-school qualification offer remote work against 23.4% of those asking for a bachelor's degree, and the experience profile rises from 5.6% below one year to 31.8% at seven to nine years before easing at ten or more.

Figure B.4 sets our measure against the occupation-level feasibility classification of Dingel and Neiman (2020), pooling U.S. postings from 2023 to 2025 and matching occupations as in Figure 4. Panel A sorts O*NET occupations into quintiles of their own remote-work share, and the share of postings in teleworkable occupations climbs from 3.7% in the bottom quintile to 95.0% in the top. Panel B repeats the comparison across the 75,590 firms with at least 100 postings and recovers a fitted slope of only 0.28 with an R^2 of 0.20.

Table B.3 asks whether the variation left over in Panel B travels with the employer. It compares firms with at least 50 postings in a focal occupation that advertise some remote work there against firms that advertise none, measuring each group's remote-work share in the same firm's *other* occupations. In occupations classified as not teleworkable, the first group advertises remote work in 13.0% of its other postings against 1.2% for firms that advertise none and 3.3% across all U.S. postings in such occupations. Firm identity alone explains far more of the variation in these cell-level remote-work shares (adjusted R^2 of 0.69) than occupation identity alone (0.24). Remote work is portable within the employer, as one expects of a policy set by the organisation rather than by the job.

Table B.4 reports country-level shares with no weighting, under each country's own 2019 vacancy composition, and under the 2024 U.S. weights behind our headline series. The choice matters most for the United Kingdom, whose 2025 share is 16.1% unweighted, 14.9%

under the common U.S. mix and 19.1% under its own pre-pandemic mix, against 9.5, 9.9 and 11.8% for the United States. The ranking of countries and the shape of every series survive the choice of scheme.

Vacancy postings and household surveys deliver different things. Our measure covers the flow of new postings monthly and in near real time from 2014 across roughly 78,000 cities in five countries and some 13.4 million named employers. The American Community Survey measures the stock of home-based workers annually for U.S. metropolitan areas with a nine-month lag, and the Current Population Survey separates hybrid from fully remote work only for U.S. states and large metropolitan areas since October 2022. Neither source dominates the other. The two can diverge even where they overlap, since postings record new hiring commitments while surveys record current arrangements, advertised terms may be renegotiated after matching, and not every vacancy converts into a hire. Across the 44 metropolitan areas with matched postings in every occupation cell of the 2024 ACS comparison, the standard deviation of log remote-work shares is 1.02 in postings against 0.68 in the survey. The vacancy data reveal substantially more geographic dispersion, part of which is sampling noise in smaller metros and part the cross-employer heterogeneity that a household survey necessarily averages out.

Geography-level results

Figure B.5 plots each city's 2025 remote-work share against its 2019 share on log scales, for the 578 cities with more than 10,000 postings in both years. The unweighted OLS fit of $\log(y) = 1.88 + 0.36\ \log(x)$ yields an R^2 of only 0.13, against 0.61 for the same exercise across occupation groups. London, Toronto and Sydney sit well above the fitted line while U.S. cities such as Memphis and Savannah sit below it, so pre-pandemic adoption only weakly anchors where cities stand today.

Firm-level results

Figure B.6 sorts U.S. postings into bins by the employer's posting volume in the year. In 2019 the relationship between remote-work shares and employer size is essentially flat, and if anything mildly negative. By 2022 a clear positive gradient has opened up, rising from roughly 8% to 11-12% at employers posting a hundred or more vacancies a year. The post-pandemic shift has been led by larger recruiters, consistent with fixed costs to redesigning jobs around remote work.

Finally, Table B.5 benchmarks our firm-level measure against the Flex Index, which assigns each employer to one of three categories running from fully on site through structured hybrid to fully flexible and which we match to Lightcast employers by name. The table summarises the 2025 posting shares of the 6,426 matched firms with at least 25 U.S. postings. Mean shares rise monotonically across the categories from 6.9 to 21.6 to 37.5%, and the interquartile range widens in step from 5.2 to 28.9 to 63.0 percentage points. Two instruments built from entirely different inputs rank employers the same way.

C Supplementary Information on Measurement

This appendix documents how RWO is estimated, how each of the alternative classifiers we benchmark it against is implemented, and the model versions, settings and prices behind Table 4.

C.1 Estimation details for RWO

RWO builds from DistilBERT (Sanh et al., 2020), which has a Transformer architecture with six layers and 66 million parameters. It was originally estimated to predict randomly deleted words in a corpus of unpublished books and all English Wikipedia. We use the uncased version of the model.

The first estimation step in RWO is to pre-train off-the-shelf DistilBERT (Sanh et al., 2020) to predict randomly deleted words in a random sample of 900,000 job posting sequences which is balanced across all years and countries. The total fraction of deleted words is 15%. We use guidelines from the original BERT paper (Devlin et al., 2019) to select the hyperparameters for estimation: a batch size of 8, three training epochs, and a learning rate of 5e-5.[21]

The second estimation step in RWO is to fine-tune the model to predict human labels. To select the estimation hyperparameters, we use three-fold cross validation and the training data used for the benchmark exercises reported in section 3. We perform an exhaustive search over learning rates $\{2 * 10^{-5}, 3 * 10^{-5}, 5 * 10^{-5}\}$, epochs $\{2, 3, 5\}$ and batch sizes $\{16, 32\}$, and select the set of hyperparameters with the highest average F1 score across training data splits. The resulting choices are 5e-5, 2, and 16, respectively. The model estimated with these choices solely on the training data is used to determine the test-set performance reported in section 3. To produce output on the entire dataset, we re-estimate the model using all human labels (from both training and test sets) using the same hyperparameters and use this model to predict remote work on all sequences in the Lightcast data.

C.2 Details for other classification approaches

Section 3 compares various alternatives to RWO for classifying remote work, and here we provide additional details on these.

C.2.1 Dictionary

We implement the dictionary approach with the following steps:

1. **Preprocessing**: We lowercase all text, remove punctuation symbols (except for hyphens and apostrophes), remove numbers, and replace all sequences of white spaces with a single white space.

[21] The batch size determines how many text sequences are processed at each step in estimation. The number of epochs determines the total number of times the entire data set is passed through in estimation. The learning rate determines the speed at which the model parameters update in gradient descent.

2. **Tagging**: We search for the appearance of any of the keyword phrases from Table C.1. For phrases containing multiple words (e.g. 'work from home') we allow for any arbitrary combination of white spaces and hyphens separating the words that compose the dictionary keyword (e.g. 'work-from-home', 'work- from- home').

3. **Binary classification**: Any job posting that contains a match to any of the dictionary keywords is classified as positive.

C.2.2 Negation adjustment

Our strategy for negation adjustment follows that proposed by Shapiro et al. (2022) to capture negation in the context of sentiment analysis. For every keyword match from the dictionary within a job posting, we consider it to be negated if any of the following is true:

1. There is a negation term in any of the three words before the keyword match. The set of negation terms is displayed in Table C.2, and comes from the VADER Sentiment Analysis toolkit.

2. "no" or "not" appear in the two words after the keyword match

3. A word that contains "n't" is the immediate word after the keyword match

If a job posting is negated we then change its binary label from positive to negative.

C.2.3 Logistic regression

Our approach to logistic regression follows the approach in Adams-Prassl et al. (2020). We start by applying the same pre-processing steps used for the dictionary approach to the job postings: i) lowercase text, ii) remove punctuation (except for the hyphen), iii) remove numbers, and iv) clean white spaces. Next we split the text into individual tokens and build the document-term frequency matrix by using the 5,000 most common tokens. For each keyword in our dictionary (the phrases in Table C.1) that is not part of the 5,000 most common tokens, we add a column in the document-term matrix with its counts. Finally, we transform the matrix into its binary form; every entry above one is replaced with a one. This matrix then becomes the set of covariates used to predict human labels via logistic regression with L_1 regularization (LASSO). To determine the LASSO penalty, we use five-fold cross-validation on the training data, and select the regularization parameter that achieves the highest average $F1$ score across the five splits.

C.2.4 Logistic regression with negation

We follow an identical procedure to the one described for logistic regression but we further extend the document-term matrix with one extra column per keyword in the dictionary that indicates that the keyword was negated (according to our negation adjustment described before).

C.2.5 Zero-Shot Learning

We use two non-reasoning models (GPT-4o and Claude Sonnet 4.6) as well as two reasoning models (GPT-5.5 and Claude Opus 4.8) for zero-shot learning. The prompt used for all models is common and displayed below.

```
You are a data expert working for the Bureau of Labor Statistics.
You are analysing the text of fragments of job postings and classifying them into
     one of two categories:

0. No remote work: the text doesn't offer the possibility of working any day of
    the week remotely.
1. Remote: the text mentions the possibility of working remotely at least one day
     per week.

You always need to provide a classification. Please provide the classification in
     following format:

-   Classification: [0 or 1]


Now classify the following text. Respond with ONLY the classification label,
    nothing else.

Text: {job-posting fragment}

Classification:
```

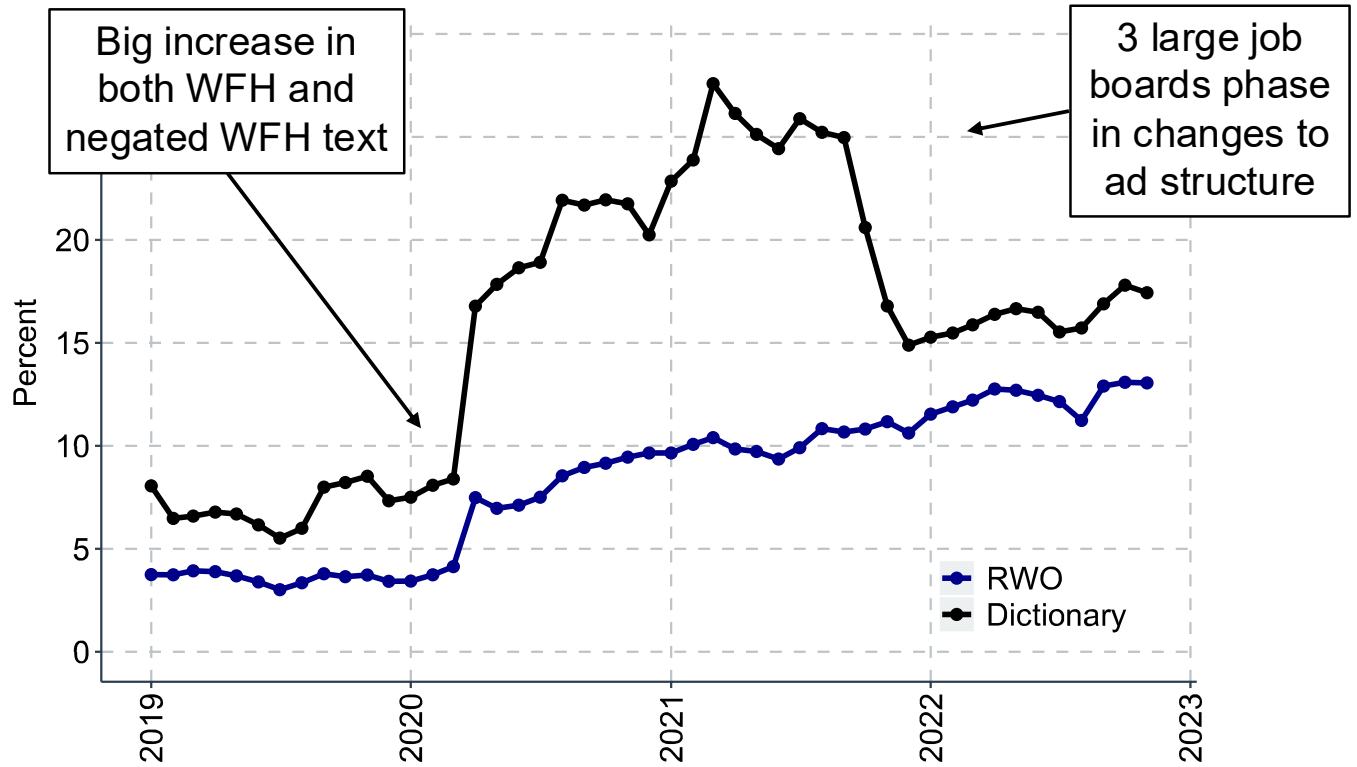
Prompt for Zero-Shot Learning

C.3 Model snapshots and settings

Frontier models are versioned and their endpoints are periodically retired, so the zero-shot results in Table 4 are reproducible only against the particular model versions we query. GPT-4o, GPT-5.5, Claude Sonnet 4.6 and Claude Opus 4.8 were each accessed through their public APIs in early July 2026. All four receive the same prompt, with no labelled examples, and all are scored on the same 4,050 held-out sequences as every other method in Table 4. We set a temperature of zero for GPT-4o; the endpoints of the other three no longer accept a temperature parameter, and for the two reasoning models we set a medium reasoning effort. RWO itself is the uncased DistilBERT checkpoint of Sanh et al. (2020), fine-tuned with the hyperparameters reported earlier in this appendix and estimated on the Google Cloud Platform using the hardware in Table B.2. The final column of Table 4 prices the classification of one million sequences as of 8 July 2026: $2.80 for RWO, $283 for GPT-4o, $436 for Claude Sonnet 4.6, $994 for Claude Opus 4.8 and $1,350 for GPT-5.5. For

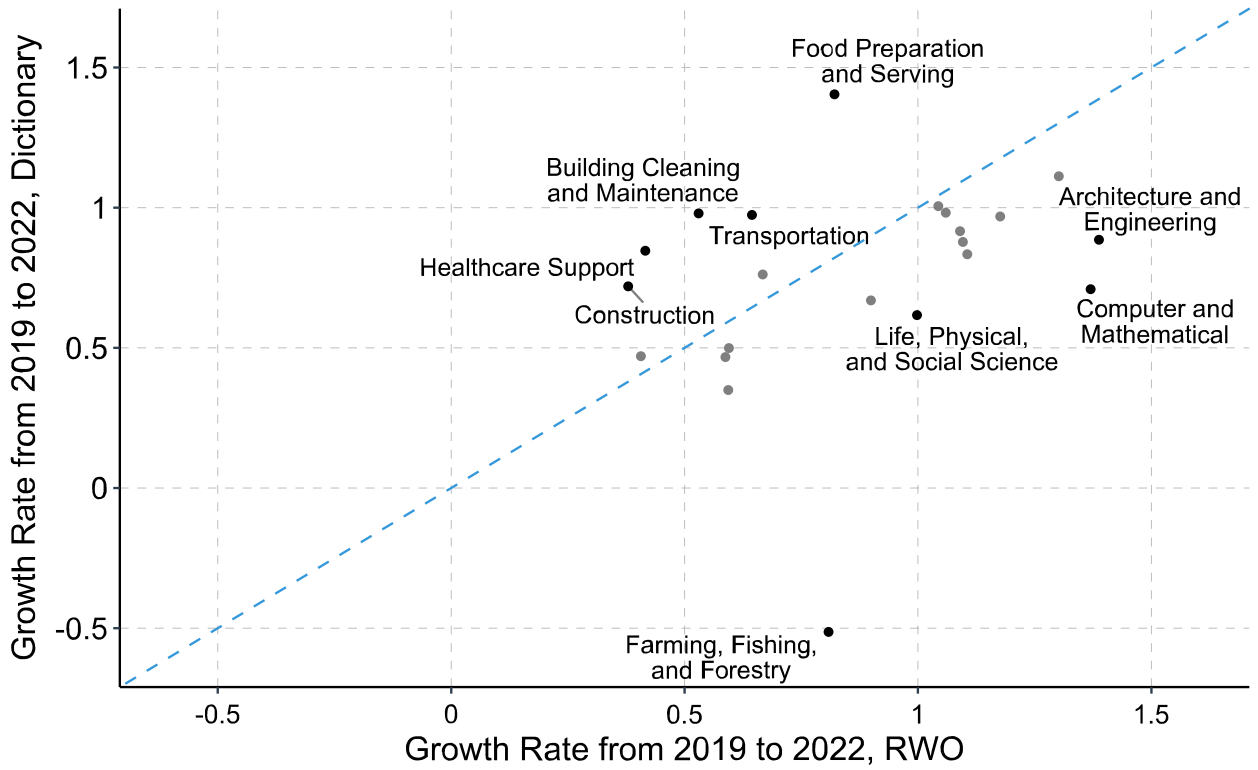
RWO this is the variable cost of the GPU, CPU, memory and disk resources consumed on a Google Cloud server; for the other four it is the published API price applied to the token volume our sequences imply.

Figure A1: RWO and Dictionary Methods Applied to U.S. Vacancy Postings



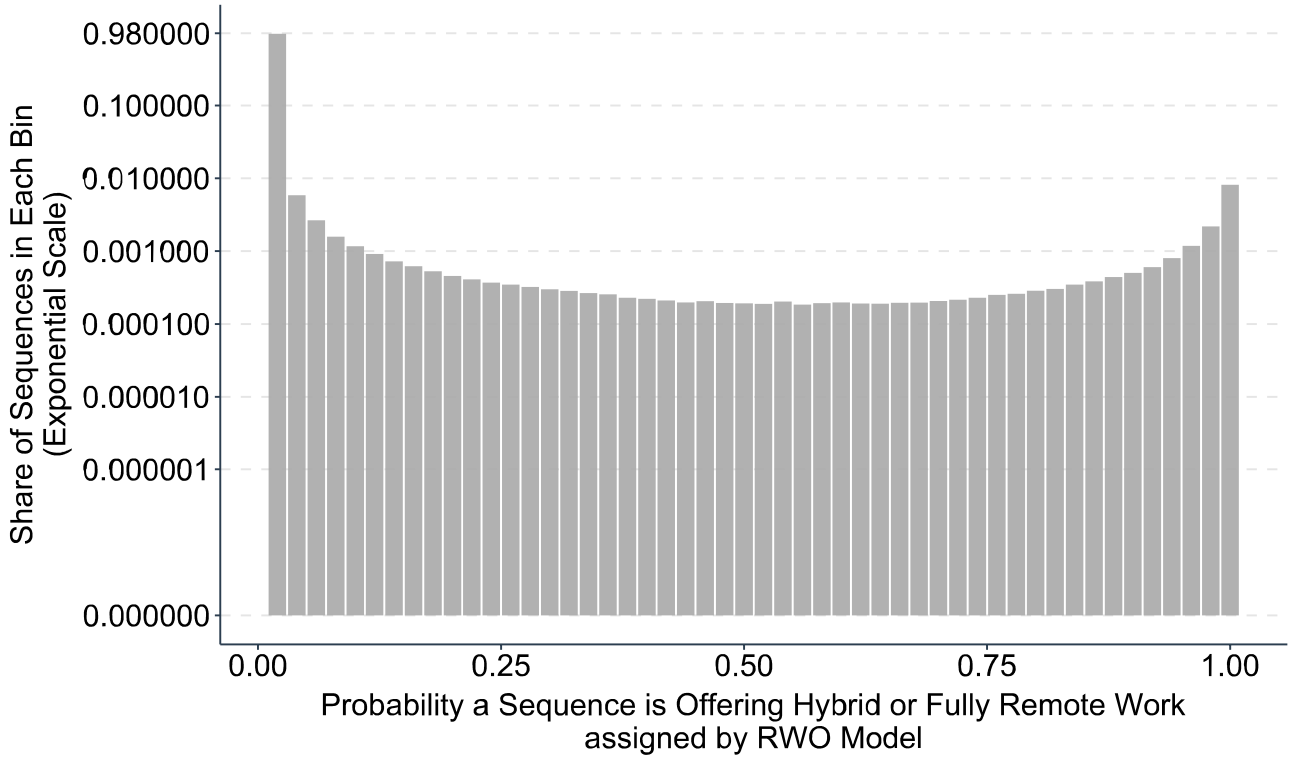
Note: This figure shows the percent of postings that say the job allows one or more remote workdays per week, as classified by our 'Remote Work in job Openings' (RWO) large language model (blue) and a dictionary-based approach (black) using the keywords in Adrjan et al. (2021). For both methods, we reweight the data to match the U.S. occupational distribution of vacancies in 2019 at the six-digit SOC level.

Figure A2: Share of U.S. Postings that Allow Some Remote Work, Growth Rate by Two-Digit Occupations, RWO Compared to Dictionary Method



Note: We sort postings into Standard Occupational Classifications (SOC) at the two-digit level and calculate the share of postings that say the job allows for one or more days per week of remote work in 2019 and 2022. We then calculate the DHS growth rate from 2019 to 2022 as (X_2022 - X_2019) / (0.5 * (X_2019 + X_2022)). For the dictionary method, we use the keywords in Adrjan et al. (2021). The blue-dashed line shows a 45 degree line.

Figure B.1: Most Sequences are Assigned a Predicted Probability by RWO at Extreme Values



Note: The 'Remote Work in job Openings' (RWO) large language model assigns a predicted probability to each sequence in the full job posting dataset using our trained neural network model. This figure presents a histogram of the share of sequences that fall in different bins according to these predictions.

Table B.1: Most Job Postings Either Have Zero or One Sequence that gets Classified as Offering Hybrid or Fully Remote Work Arrangements

(1) Remote Work Sequences	(2) Number of Vacancy Postings	(3) Share Of Total (%)
0	40,442,174	91.0
1	2,534,285	5.7
2	936,736	2.1
3	326,485	0.7
More than 3	187,183	0.4

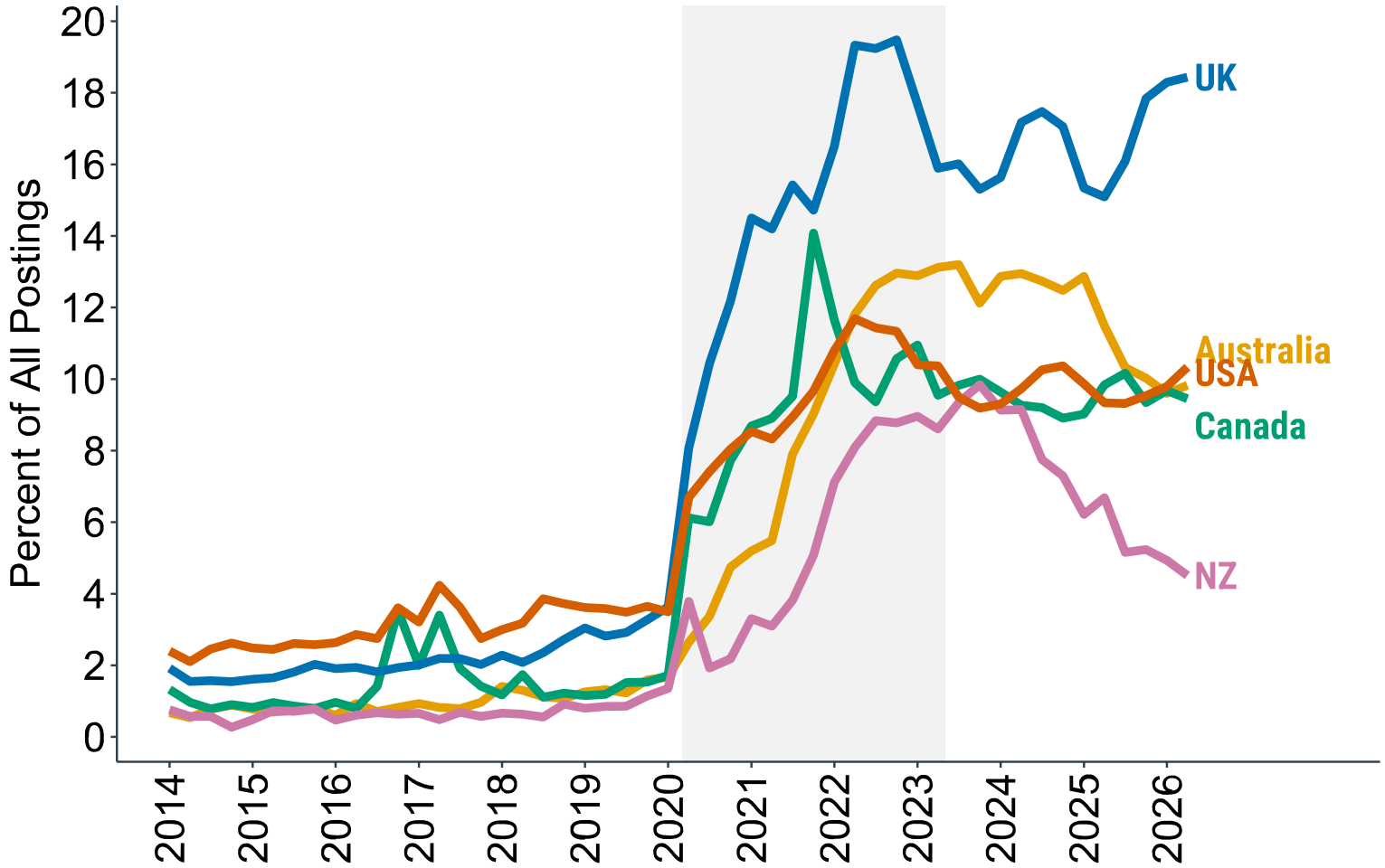
Note: This table tabulates how many text sequences in each US job posting from 2021 are classified as offering remote work (either hybrid or fully remote) according to our 'Remote Work in job Openings' (RWO) large language model. A typical job ad is split into six sequences. Most postings (91.0%) have no positive sequences. Of the remaining fraction, most have only one positive sequence.

Table B.2: Computing Hardware and Costs for the Two Training Stages of RWO

	(1) Pretraining	(2) Fine-Tuning
Computational setup	GCP Virtual Machine with 1 NVIDIA V100 GPU	GCP Virtual Machine with 1 NVIDIA V100 GPU
Total time (hours)	12	3
Cost per hour (USD)	$3	$3
Total Cost (USD)	$36	$9

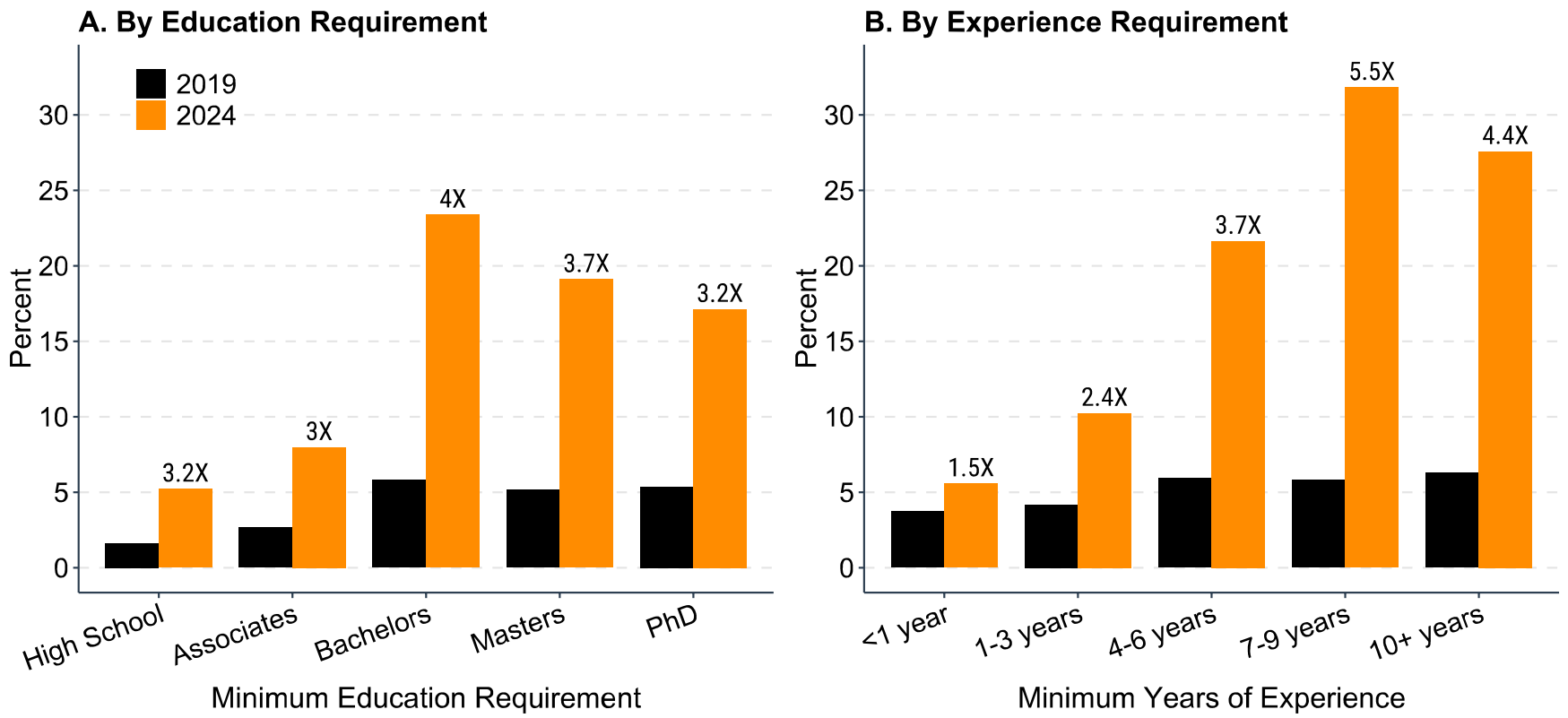
Note: This table details the computational setup and time/money costs associated with the different stages of implementing our 'Remote Work in job Openings' (RWO) large language model. All computations were performed on the Google Cloud Platform. Column (1) reports the costs associated with "pre-training" i.e. conducting additional unsupervised training of the DistilBERT model using online job vacancy posting text, Column (2) reports costs for "fine-tuning" i.e. explicitly training the model using our 30,000 human labels which identify remote work arrangements.

Figure B.2: The raw unweighted share of new job ads offering hybrid or fully remote work is highest in the UK, as UK has very high proportion of 'white-collar' jobs being advertised



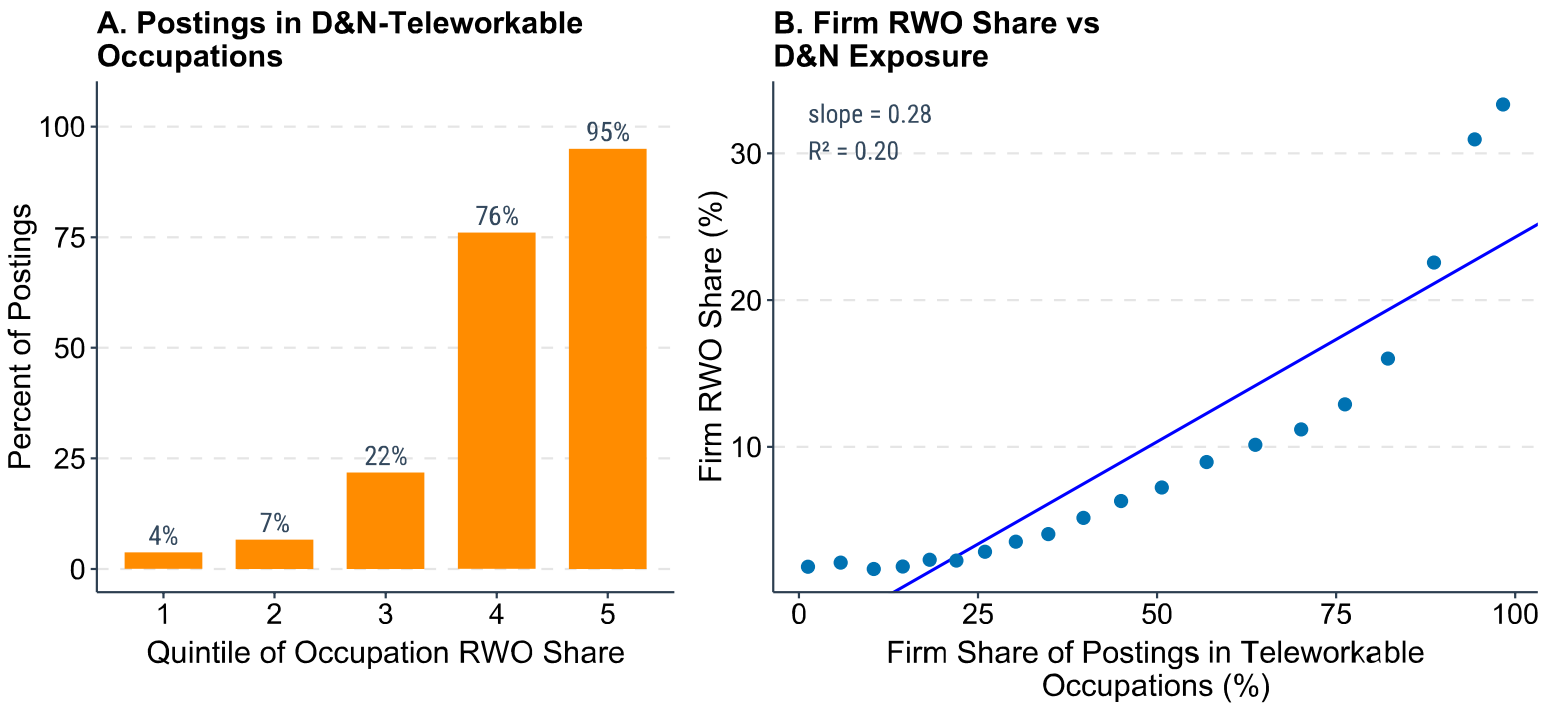
Note: This figure shows the share of vacancy postings that say the job allows one or more remote workdays per week. We compute these monthly, country-level shares as the raw mean from the universe of new job vacancy postings in each country from each month. Our baseline approach presented in Figure 2 uses vacancy shares from the US to control for occupational composition across countries.

Figure B.3: The Share of Postings that Offer Hybrid or Fully Remote Work Rises Steeply with the Required Level of Education and Experience



Note: Each bar reports the percent of U.S. vacancy postings that say the job allows one or more remote workdays per week in the indicated year and group. Panel A groups postings by the minimum education requirement stated in the posting; Panel B by the minimum years of experience required. Postings with no stated requirement are excluded. Labels report the ratio of the 2024 share to the 2019 share.

Figure B.4: Task Feasibility Only Partially Predicts Firms' Remote Work



Note: U.S. postings, 2023-2025. Panel A groups O*NET occupations with at least 100 postings in each of at least 40 states into quintiles of the occupation-level share of postings that say the job allows one or more remote workdays per week (RWO), and reports the percent of each quintile's postings in occupations that Dingel & Neiman (2020, D&N) classify as teleworkable. Panel B plots, for the 75,590 firms with at least 100 postings, the firm's RWO share against the share of its postings in teleworkable occupations (20 equal-count bins; the line is the unweighted OLS fit). Task feasibility predicts firms' remote work only partially: the fit is far from one-for-one, and firms with near-identical task exposure differ widely. Table B.3 shows that the residual variation is a firm trait that carries across occupations.

Table B.3: Firms That Advertise Remote Work in One Occupation Do So Across Their Other Hiring - Even in Jobs Classified as Not Teleworkable

Does the firm advertise remote work in the focal occupation?	Remote-work share of the firm's OTHER postings, %		
	Other occupations that are teleworkable (D&N = 1)	Other occupations that are NOT teleworkable (D&N = 0)	All other occupations
Yes	24.3	13.0	19.9
No	5.2	1.2	2.4
Memo: all U.S. postings in these occupations	19.8	3.3	9.9
Ratio, yes / no	4.7x	10.7x	8.3x
Log-odds ratio	1.77	2.50	2.32
Adjusted R^2, firm effects only	0.71	0.64	0.69
Adjusted R^2, occupation effects only	0.14	0.12	0.24
Firm-occupation pairs	162,080	167,715	182,519

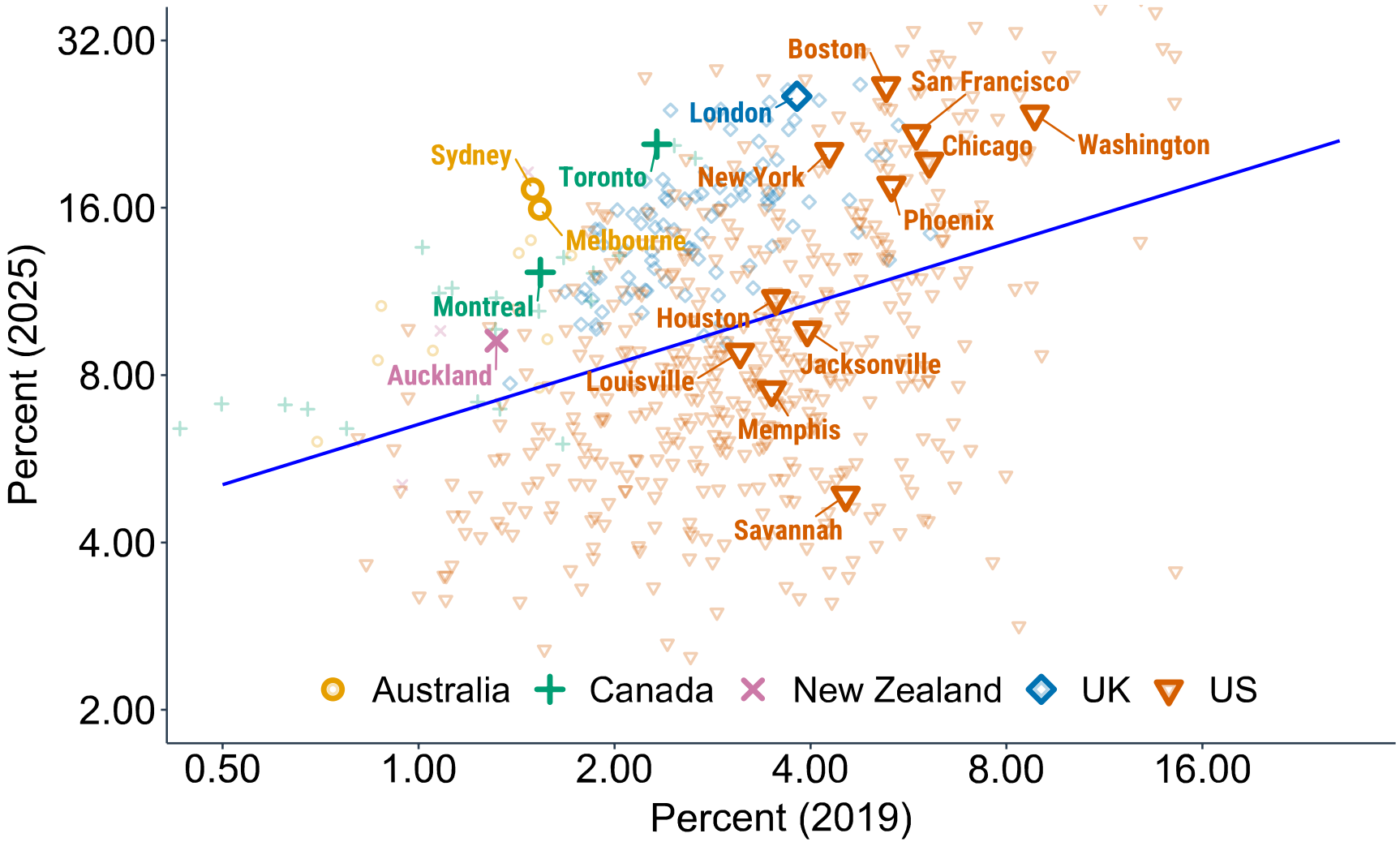
Note: U.S. postings, 2023-2025. The unit is a firm-occupation pair: a firm with at least 50 postings in a focal occupation and at least 100 postings across its other occupations. Rows split pairs on the extensive margin - whether the firm advertises any remote work at all in the focal occupation. Columns report the remote-work share of that same firm's postings in its OTHER occupations, split by whether Dingel & Neiman (2020) classify those other occupations as teleworkable. Every entry is an UNWEIGHTED mean across pairs - each pair contributes its own share equally, so large employers count no more than small ones (a firm qualifying in several focal occupations contributes one pair for each); the ratio and log-odds ratio are computed on that same basis. The memo row is the benchmark: across all U.S. postings in 2023-2025, 3.3% of those in not-teleworkable occupations advertise remote work (19.8% of teleworkable ones); it is posting-weighted over a different population - every posting in those occupations, not just these firms'. The contrast is starkest where task content says remote work should be impossible: among occupations D&N classify as not teleworkable, firms that advertise remote work in the focal occupation do so in 13.0% of their other postings - four times the national norm for such work, and more than ten times the 1.2% of firms that advertise none, which sit at about a third of the norm. Qualifying pairs average 7.7% in not-teleworkable occupations, above the national rate because the sample is restricted to larger employers hiring across several occupations. The final two rows report the adjusted R-squared from regressing the remote-work share of each firm-occupation pair on firm fixed effects alone and on occupation fixed effects alone, classifying each pair by its own D&N status; the final column pools all pairs. The adjusted R-squared regressions use 87,595 teleworkable, 90,774 non-teleworkable and 182,519 pooled firm-occupation cells.

Table B.4: Country-Level Remote-Work Shares under Alternative Occupation Weights

	2019			2025		
	Raw	**US-2024 mix**	**Own-2019 mix**	**Raw**	**US-2024 mix**	**Own-2019 mix**
USA	3.6	3.2	3.6	9.5	9.9	11.8
UK	3.0	2.4	2.9	16.1	14.9	19.1
Canada	1.4	1.3	1.3	9.6	9.0	9.5
Australia	1.4	1.1	1.4	11.2	10.0	13.9
New Zealand	0.9	0.8	0.9	5.8	5.9	7.0

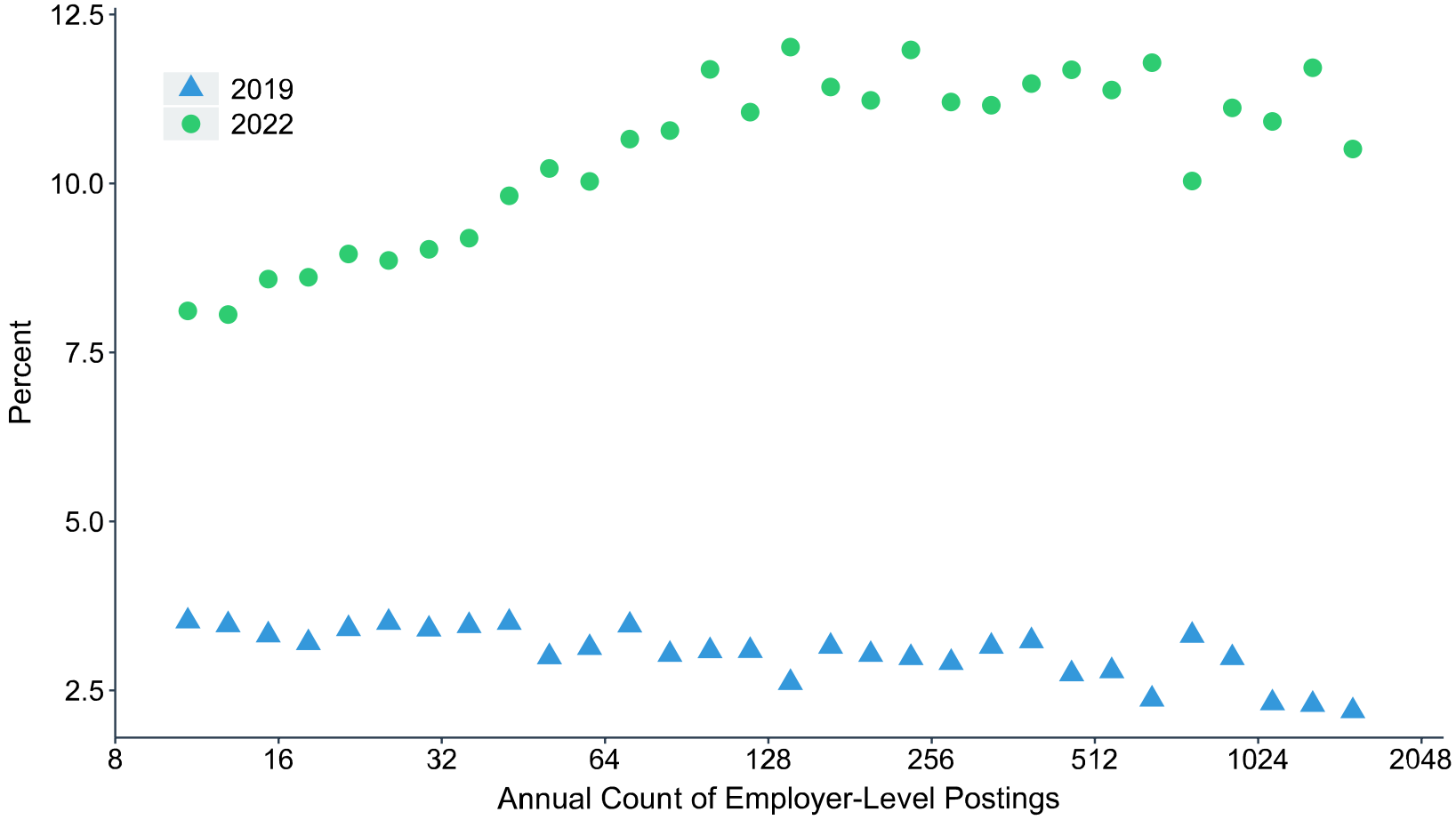
Note: Yearly means of the monthly country-level shares of postings offering one or more remote workdays per week (percent), under three occupation-weighting schemes. Raw: unweighted. US-2024 mix: the paper's baseline, weighting occupation cells by the 2024 U.S. vacancy distribution. Own-2019 mix: each country's own 2019 vacancy composition. The adjustment is largest for the UK, whose vacancy mix skews toward white-collar occupations; time-series shapes are essentially unchanged across schemes.

Figure B.5: The Share of Vacancy Postings that Explicitly Offer Hybrid or Fully Remote Work Grew at Different Rates across Cities since the Pandemic



Note: This figure plots the city-level percent of postings that say the job allows one or more remote workdays per week in 2019 and 2025. "City" refers to the location of the establishment or firm that is hiring. The line shows the unweighted OLS fit: log(y) = 1.88 + 0.36 log(x), which has an R^2 value of 0.13, across 578 cities with at least 10,000 postings in both years. Country-specific fits (slope; R^2; cities): US 0.51; 0.22; 446. UK 0.53; 0.34; 93. Canada 0.54; 0.62; 24. Australia 0.69; 0.41; 11. New Zealand 2.39; 0.81; 4.

Figure B.6: How the Share of Postings that Advertise Hybrid or Fully Remote Work Varies with the Annual Count of Employer-Level Postings



Note: For each employer, we first tabulate the number of postings in the indicated year. We then drop postings at employers with fewer than 10 postings in the year and postings by the largest 1% of employers (as measured by postings in the year). Finally, we sort the remaining postings into bins defined by the number of employer-level postings in the indicated year.

Table B.5: Firm-Level RWO Shares Align with the Flex Index

Flex Index category	Firms	Mean RWO share (%)	Median RWO share (%)	Interquartile range (IQR)
Fully On Site	2,051	6.9	1.0	5.2
Structured Hybrid	3,035	21.6	9.5	28.9
Fully Flexible	1,340	37.5	25.6	63.0
All matched firms	6,426	20.3	6.2	26.7

Note: Firms matched between the Flex Index (Sept-2025 snapshot) and Lightcast by name, keeping the 6,426 matched firms with at least 25 U.S. postings and a measurable remote-work share in 2025. Each cell summarises firm-level RWO shares (percent of the firm's postings offering one or more remote workdays per week); the final column is the interquartile range across firms. Flex Index categories are as reported by the provider.

Table C.1: Keywords Used to Implement the Dictionary Approach to Remote-Work Classification

working remotely	working from home	work remotely
work from home	work at home	teleworking
telework	telecommuting	telecommute
smartworking	smart working	remote work
remote	remotely	homeoffice
home office	home based	homebased

Note: These are the keywords that appear in Table A.2 of Adrjan et al. (2021) for detecting the presence of remote work in the text of job postings. The three exceptions are `homebased', `home based', and `remotely' which we add to the original terms to improve accuracy.

Table C.2: Terms Used for Negation in the Dictionary Approach

aint	arent	cannot	cant	couldnt	darent	didnt	doesnt
ain't	aren't	can't	couldn't	daren't	didn't	doesn't	dont
hadnt	hasnt	havent	isnt	mightnt	mustnt	neither	don't
hadn't	hasn't	haven't	isn't	mightn't	mustn't	neednt	needn't
never	none	nope	nor	not	nothing	nowhere	oughtnt
shant	shouldnt	uhuh	wasnt	werent	oughtn't	shan't	shouldn't
uh-uh	wasn't	weren't	without	wont	wouldnt	won't	wouldn't
rarely	seldom	despite	no				

Note: This is a set of terms that the VADER sentiment analysis tool uses for negation, and which Shapiro et al. (2022) adopt. We add the term `no' to the baseline negation set.