Remote Work across Jobs, Companies, and Space*

Stephen Hansen†, Peter John Lambert†, Nick Bloom‡, Steven J. Davis§,
Yabra Muvdi*, Raffaella Sadun**, Bledi Taska‖

Economics Working Paper 23104

HOOVER INSTITUTION
434 GALVEZ MALL
STANFORD UNIVERSITY
STANFORD, CA 94305-6010
January 13, 2026

The pandemic catalyzed an enduring shift to remote work. To measure this shift, we examine more than 500 million job vacancy postings across five English-speaking countries. Our measurements rely on a large language model (LLM) that we fine-tune with 30,000 human classifications. The model achieves 99% classification accuracy, substantially outperforming dictionary-based approaches and—despite being a fraction of the size—performs on par with frontier AI models. From 2019 to 2025, the share of postings indicating that 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 zooming in 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. Data accompanying this paper can be found at www.WFHmap.com.
† University College London
† London School of Economics & University of Warwick
‡ Stanford University
§ Hoover Institution
* Digitec Galaxus AG
** Harvard Business School
‖ 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 subsided. 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 or else assessments of remote-work feasibility by occupation. We rely instead on the information contained in job vacancy postings. Specifically, we consider the full text of over 500 million postings in five English-speaking countries. In doing so, we build and deploy a large language model (LLM) to analyze the text and determine whether the job allows for remote work. We fit, test, and refine our language-processing model using 30,000 classifications generated by human readings. To the best of our knowledge, this research effort was one of the first to implement such a large-scale application of large language model for economic measurement.[1]

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 the currently employed. Indeed, the remote-work share of postings lags far behind the remote-work share of employment in the pandemic's early stages. 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.

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 2014 through late 2025 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.

[1]We publicly released our open-source code and accompanying data in March 2023, only a few months after the debut of ChatGPT. Since that time, the use of LLMs across society—including for economic measurement—has grown rapidly. Despite the dramatic advances in the capability and scale of available models, our own benchmarks indicate that the methodology presented in this paper remains state-of-the-art and comparable in performance to much more recent generic AI models for this specific application, while being deployable at only a small fraction of the computational resources and cost required by frontier systems.

Comparable detail is hard to obtain from other sources, especially at scale.[2] Third, we apply frontier methods to develop a language-processing model that reads and classifies postings in an automated manner. The model achieves a 99% accuracy rate in flagging jobs that allow for remote work, greatly outperforming dictionary methods. Our model also outperforms a variety of other methods, including frontier foundation AI models, while being far less costly to deploy at scale. Fourth, we validate our vacancy-based measure against official Census Bureau benchmarks from the American Community Survey and Current Population Survey, finding strong correlations (0.70–0.93) across metropolitan areas and occupations.

The combination of scale, rich text data, and automation lets us characterize the shift to remote work in a highly granular manner. The near-universal coverage of vacancy postings, totaling millions per month, allows us to trace the diffusion of remote work across occupations, industries, locations, and employers. Few alternative data sources can match this scale and level of detail without substantial cost, making it possible to explore new questions about the extent, origins, and implications of remote work.

We post many statistics at www.WFHmap.com, with frequent updates, which have been viewed and downloaded by 10,000 unique visitors a year since launch. We make additional data available on request, and have done so hundreds of times including for dozens of policy institutions. The data also informs many academic papers, including Autor et al. (2023), Barrios et al. (2024), Cowan and Garcia (2024), and Bloom et al. (2026) It has also featured in two Economic Reports of the President (Council of Economic Advisers, 2024; Council of Economic Advisers, 2025).

The share of postings that say new employees can work remotely one or more days per week was small before the pandemic: between 0.8% and 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. By late 2025, the remote-work share stands at approximately 9-10% in Australia, Canada, and the U.S., and nearly 16% in the U.K.—a "new normal" that shows no meaningful sign of reverting to pre-pandemic levels. New Zealand is a partial exception, showing some reversion from its 2024 peak while remaining well above pre-pandemic levels. These patterns continue to evolve, underscoring the value of ongoing measurement as employers and workers adjust to shifting norms around workplace flexibility.

[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).

Remote-work posting shares vary greatly across occupations and 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. According to a linear least-squares regression, 80% of the variation across occupations in 2022 remote-work shares is accounted for by their 2019 shares. In contrast, just 11% of city-level variation in 2022 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 when they are recruiting in the same occupational category. This emergent heterogeneity on the demand side expands opportunities to satisfy preferences over remote work on the supply side.[3] Our non-uniformity result also carries 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.

We use a relatively small large language model—DistilBERT[4] (Sanh et al., 2020)—to measure remote work. First, we pre-train DistilBERT on one million text chunks drawn from vacancy postings. This step familiarizes DistilBERT with the (heterogeneous) structure of job ads. Second, we consider 10,000 text chunks 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 as follows: the job allows hybrid or fully-remote work arrangements, or it does not. Finally, we apply the resulting "Remote Work in job Openings" (RWO) large language model to classify all 500 million job postings.

The RWO model greatly outperforms a previously used dictionary in classifying the

[3]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).

[4]DistilBERT is a smaller, faster version of the BERT model introduced by Devlin et al. (2019). BERT and DistilBERT exploit machine-learning tools and are pre-trained on the full English-language Wikipedia corpus and the Toronto Book Corpus. For a helpful non-technical overview of BERT, see Luktevich (2022).

remote-work status of vacancy postings.[5] The dictionary method yields high classification error rates that vary greatly over time and across occupations. Expressions like "home or office working possible" and "work from home care facilities" and "requires a Home Office work permit" suggest some of the difficulties that arise when applying dictionary methods to job ads. Logistic regressions and generic AI models offer large improvements over dictionary methods. Our RWO model classifications offer even larger improvements, outperforming all other methods when applied to vacancy postings, as measured by accuracy rates, precision, and F1 scores.

Relatively few works in economics combine a deep learning model with human-generated labels to develop an automated classification model and to quantify its performance. On the other hand, with the advent of generic AI models, there is a growing trend to using zero- or few-shot learning for classification (see Ash and Hansen (2023) and Ash et al. (2025) for recent reviews). An important methodological point of our paper is that smaller, fine-tuned models can outperform large models while also working within a fully open-source environment.[6]

A prominent line of research classifies occupations as suitable or unsuitable for remote work based on descriptions of work activities and experiences.[7] 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 2024 ranges from 0 to 57% with a mean of 4% and standard deviation of 7%. For occupations they classify as suitable for work from home, the share ranges from 0.1 to 58% with a mean of 16% and standard deviation of 12%.

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

[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).

[6]Shapiro et al., 2022 develops a BERT-based model and finds little gain relative to dictionaries in detecting news sentiment. However, it uses fewer than 1,000 human-labeled text examples in fitting a BERT-based model, which may explain why it yields small performance gains. 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]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.

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.

2 Data and Measurement

To measure remote-work posting shares, we exploit a near-universe of online job postings from January 2014 through November 2025 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 (Sanh et al., 2020) as the foundation model for the remote work classifier. DistilBERT begins from the BERT language model which is pre-trained on thousands of books and the English-language Wikipedia corpus.[8] 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) large lan-

[8]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.

guage 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 incluide 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. See Appendix A for a detailed description of our data and pre-processing steps.

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 online 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, location of the employer, industry, and more. We consider postings listed from January 2014 and November 2025, 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 more than 550 million online vacancy postings in five countries, spanning 12.5 million employers and nearly 62 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 online vacancy postings in 2024. Appendix B reports selected results for alternative weighting schemes.

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[9] 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 binding 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 size of our data, however, this approach is not feasible to scale and some means of automated classification is required. A traditional approach adopted in the text-as-data literature in economics is to use a dictionary of keywords whose presence is assumed to indicate 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).

An issue that becomes immediately apparent upon inspecting job postings that are classified by keywords is the presence of notable errors, which Table 2 illustrates. False positives include references to companies' home offices and working in homes dedicated to healthcare provision. A second, and perhaps most worrying, source of false positives is that the structure of job ads shifts during COVID-19 in a way correlated with the presence of false positives. This is due to the fact that, after early 2020, many postings feature a new text field indicating whether home work is allowed, and then explicitly state it is not—a naive application of the dictionary method would infer from this text field that the job posting allows working from home.[10] Table 2 also lists examples of false negatives, which illustrates the many and complex ways that companies can use to describe remote work. Accounting for this linguistic variety with a fixed set of keywords is a major challenge.

2.3 Our Approach to Classification

Our approach to address the classification errors in the dictionary approach has three steps. First, we use three human auditors to read and classify 10,000 pieces of text extracted from

[9] In principle our measurement approach could be extended to the intensive margin (days per week), but for simplicity we begin with the this binary classification.

[10] One approach to correcting this problem is to extend the dictionary to incorporate negation (e.g. to treat as a negative classification the phrase 'this is not a remote work position'). In section 3 we show that this indeed improves measurement accuracy but not by as much as our proposed solution below.

job ads which produces 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 hope is to scale the accuracy of human reading—which can only be deployed on a small fraction of data—to the entire dataset. While this approach is common in the machine learning literature, it is not often used in economics, even though it appears to hold great promise. We call the final classifier used in this paper the "Remote Work in job Openings" (RWO) large language model, or simply *RWO*.

The main text contains a broad overview, with further 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 the skills required for the job, the tasks involved, etc. Mixing text relevant for work-from-home with a great deal of irrelevant text introduces noise into the classification algorithm.

The 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." We extract these as a single sequence. Second, the beginning of each posting typically has a number of bullet points or other structured fields. In most cases, these also form a single sequence.[11] Finally, the remainder of a posting is typically structured like standard prose with a succession of paragraphs. Each paragraph is taken as a single sequence, unless it passes a length threshold. In this case, we break it into multiple sequences of consecutive sentences.

This procedure produces approximately 1.6 billion sequences out of the over 500 million job postings.

2.3.2 Human Labels for Training and Evaluation

From the sample of sequences, we first chose 10,000 to label manually. One quarter of these sequences was chosen 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 work-from-home language, including the generic terms 'remote', 'home', 'work', 'location'; any word that begins with 'tele'; and any two-word sequence that begins

[11] The exception is if the number of distinct structured fields is too large, in which case we split them into multiple sequences.

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 year-quarter from 2014Q1 through 2021Q3. We also balance the sample evenly across countries[12] to account for varying English idioms in different geographic locations.

We used Amazon Mechanical Turk to generate labels. To ensure high-quality workers, we set up an initial screening test that required prospective workers to label 20 sequences that we had previously manually classified. Only workers that made at most one error were allowed to proceed to label the full set. Another quality control strategy was to pay around 25% above typical market rates for labeling tasks. This motivated workers who passed initial screening to continue on the project.[13]

Each of the 10,000 sequences were labeled by three distinct workers. There is a high agreement rate among workers: 66.9% of examples are unanimous negative examples and 25.5% are unanimous positive examples. The remaining 7.6% examples are evenly balanced between one dissenting vote for either positive or negative. Note that, while half of the sample was chosen to contain a word with the potential to denote work-from-home, only 29.2% of the sequences receive a majority of positive votes.

2.3.3 Developing the 'Remote Work in Job Openings' (RWO) model

The field of natural language processing has been revolutionized by models that allow the meaning of word sequences to arise by how they interact. Consider the 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 the word 'remote' but only the latter is a positive example of remote work. The important point is that the interaction of 'remote' with surrounding context words determines the overall 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

[12] We draw one quarter of this dataset from each of USA, UK, Canada, and a further one quarter from the pooled Australia and New Zealand data.

[13] In general workers appeared engaged and focused on the labeling task. We received communication from multiple workers seeking to clarify ambiguous cases, which went above and beyond what AMT required for payment.

should be "paid attention to" in resolving these meanings. Self-attention is the key idea that powers all widely-used Large Language Models. See Ash and Hansen (2023) and Ash et al. (2025) for further details. DistilBERT is one such model.

We make two main modifications to the off-the-shelf DistilBERT model to build *RWO*. First, the initial set of parameters of off-the-shelf DistilBERT is 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 which 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.[14] Appendix C details how we specify the prediction model's hyper-parameters. Table 3 provides an illustration of which words are influential 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. In the following section, we compare the test-set accuracy of the estimated model with that of other algorithms in the literature and show its performance is outstanding.

2.3.4 Predicting Remote Work Language at Scale

Finally, we use the estimated 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 B1 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 work-from-home language because, as we show below, most job postings do not explicitly mention the possibility to work from home 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 typically produces a clear prediction, in line with human labelers' high agreement rates. We use an 0.5 threshold

[14]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.

for assigning a sequence a positive classification according to RWO's predicted probability, but the properties of the predicted probability distribution imply 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 job-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 sequence), the majority have just a single positive sequence. This reduces concern that the algorithm produces correlated false positive hits at the posting level.[15] 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

To allow researchers to 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`, which will reproduce the same probabilities as in the paper.[16]

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 step of development, 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 total run time for all steps is 94 hours, and the total cost is approximately $3,200. Our view is that researchers should therefore not view computational costs as a major impediment to adopting customized, purpose built large language models.

[15] We have manually read a number of randomly drawn postings with more than five positive sequences, and found no instance of the algorithm failing. In some cases, the scraping procedure that gathers data from online job portals appears to have identified as a single job ad a succession of postings by recruitment agencies. In other words, the measurement error arises from the data itself rather than the classification approach.

[16] The model is subject to revision, at which point the predicted probabilities for a given text may change. Users who find systematic biases in the predictions are welcome to contact the authors with their findings, which can be incorporated into future work.

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 undertake a systematic comparison of the ability of different algorithms to correctly classify unseen texts. 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. By way of comparison, we also use the following alternative methods for classifying test-set observations (full details of each approach are 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),[17] and count an observation as positive if it contains a term from this set.

3. *Dictionary with Negation.* Shapiro et al. (2022) shows that accounting for negation can improve the performance of dictionaries. We adopt a similar method and only count the presence of a dictionary term as indicating remote work when a negation term does not appear in the surrounding context.

4. *Logistic Regression.* Adams-Prassl et al. (2022) uses Lightcast data from the UK to measure the prevalence of flexible work schedules, i.e. the times at which work must be completed, from job posting text. The paper uses humans to manually annotate 7,000 texts, and fits a (penalized) logistic regression model for classification. The features of the logistic regression are the word frequencies in a given document. 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.

[17] The terms are reported in Table 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.

6. *Zero-shot learning with generic AI models.* Since at least the introduction of GPT-3, researchers have observed emergent behavior in LLMs that allows them to correctly answer questions outside their training objective (Brown et al., 2020). This is known as zero-shot learning. We use two recent AI models (GPT-4o and Claude Sonnet 4) to query for the presence of remote work.

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. A straightforward metric is the error rate, i.e. the fraction of mis-classified texts. On this measure, RWO outperforms all other methods with an error rate of 0.02.[18] The nearest competitors are GPT-4o and RWO without unsupervised fine-tuning (error rates of 0.03), while the dictionary method's error rate is eight times higher.

A more standard performance metric in the machine learning literature is the F_1 score which accounts for both a classifier's ability to recover the true positives (*recall*) as well as the share of predicted positives that are true positives (*precision*). The F_1 score varies between 0 and 1, where higher values indicate better performance. Again, we observe that RWO outperforms all other measures.[19]

One concern is that the distribution of positive and negative postings in the test data does not correspond to that of the full population of job postings: the data extracted for labeling is specifically designed to over-represent positive cases. To obtain a sense of classification accuracy on the full population, we create a simulated dataset of 1000 * 4,050 = 4,050,000 observations, 3% (97%) of which are sampled with replacement from the set of positive (negative) test set examples. Table 4 reports the same metrics as Table 4 but computed on this more unbalanced dataset. 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 0.85 F_1 score, while the F_1 score of GPT-4o falls to 0.73 and other methods have even worse performance. Moreover, unsupervised fine-tuning becomes more important as the F_1 score for RWO with generic embeddings drops to 0.78. These results arise because, as Table 4 shows, RWO has a particularly low false positive (FP) rate compared to other methods. When negative

[18] 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 set error rate is 0.02 in each year except for 2015 (error rate 0.03) and 2014 (error rate 0.01).

[19] An alternative dictionary for measuring remote work adoption is proposed in Draca et al. (2022) which uses our same UK Lightcast sample. The overall error rate of this dictionary in the full test data set is 0.19 and for the test data set arising from the UK is 0.17. Interestingly, the F_1 score we obtain for logistic regression (0.81) is similar to that reported by Adams-Prassl et al. (2022) for classifying flexible work scheduling (0.83, see Table 3 of that paper).

examples dominate the evaluation sample, correctly classifying them becomes important for overall performance and RWO is strong in this dimension. Since this sample's label composition is more in line with the expected 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 relatively few large-scale exercises for benchmarking the performance of different text classification algorithms for economically relevant measurement 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) does not report large gains from using BERT over simpler models for classifying news sentiment. One reason that we, in contrast, do find large gains is the size of our training data. Shapiro et al. (2022) trains BERT on 800 labeled articles whereas we have an order of magnitude more training data, which provides more information for estimating the complex ways in which word sequences map into outcomes. We conjecture that other prediction problems using text in economics might similarly benefit from a large training sample combined with sequence embedding models.

Second, fine-tuning relatively small, open-source models like DistilBERT on domain-specific text and human labels outperforms zero-shot learning with modern AI models. There are other important advantages as well. Scaling zero-shot learning to the full dataset of over one billion sequences would be prohibitively costly. At a deeper level, the training and deployment of AI models developed in the private sector are opaque and generally not reproducible. 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 compares to alternative methods on the full data sample. Rather than consider all alternatives, we focus on how RWO compares to the Dictionary method, which is most common in the literature measuring remote work adoption from 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.[20] The patterns present in both series differ markedly. 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 a steady growth rate thereafter. Two features of

[20]These time series are computed using the approach we adopt for the baseline results discussed in the next section, and are not the simple raw positive share.

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 not to have induced a discontinuity in our RWO measure, likely because it is robust to changes in structure so long as the intended meaning remains consistent. Clearly, then, the choice of measurement approach can have important quantitative implications even in aggregate.[21]

Of course, aggregate comparisons between methods can mask underlying differences at more granular levels. To illustrate this, we compute the growth rate in remote work adoption according to the Dictionary method and RWO from 2019 to 2022 for individual SOC2 occupations, pooling all 2019 postings and 2022 postings together. In these two years, the Dictionary method appears similar to RWO but with an upward shift of around five percentage points. However, as Figure A2 reveals, there are large differences in the specific occupations that each method associated with growth in remote work adoption. According to the Dictionary method, the 'Food Preparation and Serving' occupation has experienced highest growth in adoption, while for RWO the highest-growth occupation is 'Computer and Mathematical'. Moreover, according to RWO all occupations experienced positive growth in adoption, whereas adoption rates fall for the '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 occupations is more reliable. In the next section we provide a more in-depth analysis of occupation-level heterogeneity according to RWO.

In sum, the RWO model displays a very high classification accuracy—relative to human labels—and differs markedly from the most popular alternative approach in the literature based on keyword search. This difference is especially pronounced since 2020, even at the aggregate level. We believe our approach to measurement provides a highly accurate classification of remote work offers in the text of job postings, and base the remainder of the paper on analyzing its output.

4 Results

In this section we document how the *percent of new remote work vacancies*–the fraction of all new vacancies which explicitly offer the right to work remotely one or more days per

[21] The patterns in the Dictionary series need not match those from Adrjan et al. (2021) even though we use a similar set of keywords, as the structure of the Lightcast data could differ in important ways from that of the Indeed data that Adrjan et al. (2021) use.

week–has changed over time. We document this across countries, occupations, cities, and employers. This covers both hybrid and fully remote work.

This section is organised as follows: First, we look at the percent of remote work vacancies across each of our five countries. We plot this as a monthly time series, spanning January 2014 to November 2025. Second, we compare the percent of new remote work vacancies across broad and narrowly defined occupations, contrasting our measurements in 2019 and 2024. We show that the substantial rise since the onset of COVID is highly uneven across occupations, and find that occupations with the highest 2019 percentage of remote work were the most likely to top the list in 2024. We also compare occupation-level classifications used in the literature to our measurement. Third, we compare the percentage of new vacancies offering remote work arrangements across cities. We show that cities with higher remote work percentages in 2019 do not strongly predict higher percentages by 2024 (unlike occupations). This suggests that additional confounding city-level characteristics have played an important role in the adoption of remote work. We also compare a monthly time series across a selection of US cities. Fourth, we compare our measures to survey information from the American Communities Survey (ACS) and Current Population Survey (CPS). We show that MSAs which have a high remote work share of vacancies in our data also have high fractions of the population who selected "Worked from home" when asked about their commuting methods. Fifth, we show that the percentage of remote work vacancies posted by employers who operate in the same industry, and search for the same talent, can vary widely.

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 1 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 remote work vacancies across nearly 800 narrow occupation groups. We weight each group based on the share of vacancies in this group in the USA during 2024. Our baseline results utilise this method to reduce the impact of compositional differences, both across time and across countries. Four high-level facts emerge:

1. **Unprecedented and 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 in the years 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 vacancies occurred later in 2021 as their local pandemic experiences worsened.

3. **Remarkable persistence long after the pandemic shock**
   By November 2025, long after the forcing event of the pandemic and its associated policy mandates subsided, the share of remote job postings remains near or above peak levels for most countries. This persistence is most visible in the UK, which has seen continued growth to reach nearly 16% 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 partial reversion from a 2024 peak, though still remains well above its pre-pandemic baseline.

4. **Substantial heterogeneity 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 1. Comparing the unweighted Figure B.2 to Figure 1 tells us the direction and magnitude of the impact that occupation composition plays in our results. For example, by the end of 2025 the difference between the UK and USA is approximately 8.5 percentage points using the raw data (roughly 17.5% vs 9%) and 6.5 percentage points after re-weighting (roughly 16% vs 9.5%). This reduction suggests that a meaningful portion of the difference in advertised remote work shares between the US and UK is accounted for by differences in the types of jobs being advertised, which is unsurprising as the UK employment composition is skewed towards more white-collar jobs.

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 2 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 'Computer and Mathematical' occupations explicitly offered remote work arrangements, whereas in 2024, this sector leads the sample, having grown by 4.7X. Other white-collar professions show similar expansions; for instance, 'Legal' occupations grew by 4.6X and 'Business and Financial Operations' by 3.8X. Interestingly, even sectors requiring physical presence show high relative growth from low baselines; 'Food Preparation' postings mentioning remote work multiplied by 11X, 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. Lastly, Figure 2 provides some evidence that the 2019 shares of remote work correlate with post-pandemic shares.

To investigate the relationship between 2019 and 2024 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 2024 percent of advertised remote work (on a log-scale), presented in Figure 3. After dropping a handful of data points with fewer than 250 postings in 2019 or 2024, we retain 719 O*NET occupations.[22] Figure 3 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.[23] 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. This model yields an R^2 of 0.80, indicating that the share of vacancies advertising remote work in 2019 was strongly predictive of the share in 2024. The estimated slope coefficient suggests an elasticity of 0.81%, meaning the growth has been largely proportional to the initial base.[24] 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

[22] In total, there are 867 O*NET occupations. Our sample of O*NET codes which have greater than 250 vacancy postings in 2019 & 2024 is 719. This attrition is expected, for example a number of military occupations are not present in our data.

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

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

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.[25]

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.[26] 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.

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.[27]

Figure 4 shows the percent of advertised remote work across a selection of large international cities, comparing 2019 to 2024. We see that the percentages vary widely. For example,

[25] 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.

[26] A clear example of the differences between our measurement approach and Dingel and Neiman (2020) is for teaching jobs. For example, while D&N correctly classify jobs for "kindergarten teachers" as being *feasible* for full time home working (i.e. via a virtual class room), we know anecdotally that this arrangement was very taxing on staff and avoided as soon as normal schooling resumed. We find that teaching jobs in general (and "kindergarten teachers" in particular) have some of the lowest shares of advertised remote work of any job, highlighting that feasibility and actual behaviour can vary markedly.

[27] Since the predominant remote work arrangements are hybrid, the location of the work site remains a key feature of most jobs. However, in the case of a 'fully remote' position this analysis becomes less precise. We plan to refine our measurement approach in future work to distinctly classify 'hybrid' vs 'fully remote' work arrangements, but have thus far concluded that the majority of remote work jobs offer hybrid arrangements.

in 2024, more than 1-in-4 new job postings in Washington (DC) advertised remote work arrangements, compared to roughly 1-in-10 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 vacancies, with Sydney and Melbourne growing 13.8x and 11.5x respectively from their 2019 baselines.

Further evidence of the large shift in both levels and spread of remote work offers in job ads is shown in Figure 5, which plots the relationship between the pre-pandemic baseline (2019) against the share observed in 2024, across a wider sample of cities. The axes are scaled logarithmically. The unweighted OLS fit—represented by the solid line ($log(y) = 1.98 + 0.34log(x)$)—suggests a positive correlation: cities that had higher remote work adoption prior to the pandemic generally maintain higher levels today. However, the regression shows a coefficient of determination (R^2) of only 0.11, compared to a value of 0.80 when running the same exercise across occupation groups. This highlights that while 2019 shares are predictive, there is significant deviation from the trend. As shown in the scatter plot, cities in the UK, Canada, and Australia (such as London, Toronto, and Sydney) consistently appear above the regression line, indicating that they have adopted remote work practices at a faster rate than the global trend predicted by their 2019 levels. Conversely, a cluster of US cities, including Memphis and Savannah, fall below the trend line, suggesting a slower relative expansion of remote vacancy postings in those markets.

This sizable increase in the levels and spread of remote work across cities, as well as the weak relationship between 2019 and 2024 shares, 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. We leave a more formal tests of these predictions to future work.

We next turn to more granular monthly time series for selected US cities, shown in Figure 6, where we observe data right up to November 2025. 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 Validating the RWO Measure against Census Bureau Benchmarks

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 "ground truth" benchmarks from the U.S. Census Bureau: the American Community Survey (ACS) and the Current Population Survey (CPS). Figure 7 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 7 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.[28] 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 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 7 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 7 suggests that our RWO measure of remote working opportunities constructed using the flow of new vacancy postings is a highly robust

[28] 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.

indicator of the cross-sectional incidence 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.

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 reducing 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.

Our analysis of employers is by no means exhaustive, and we leave for future work a more in-depth match to firm-level covariates. The first piece of analysis illustrates that the prevalence of employers that explicitly offer remote work arrangements in their vacancy postings varies greatly, even among same-industry firms recruiting in the same occupational category. Panel A of Figure 8 shows the share of remote work vacancies posted by four large aerospace manufacturing firms (NAICS code 3364). We consider only management occupations in this panel. The data reveals 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 12% 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 8 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. 98%) 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 8 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 35% of postings offering remote work, drastically reduced this to roughly 7% 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 32% in 2025, effectively swapping positions with Honda. Ford also saw growth, rising to roughly 14% in 2025. Tesla job postings remain an outlier, making almost no offers of remote work across all observed years.

5 Conclusion

This paper's first contribution is to develop a methodology for classifying job postings as offering fully remote or hybrid work arrangements. We take an off-the-shelf, large-scale Transformer model and adjust it to both account for the specific language structure of postings and, more importantly, to predict tens of thousands of human-labeled sequences. The resulting RWO algorithm outperforms existing methods in terms of out-of-sample classification accuracy, including recent AI models. More broadly, we view the fine-tuning of small, open-source models as a compelling tool for economic measurement problems.

With our RWO measures in toe, we next parsed the near-universe of job vacancy data across five English-speaking countries (USA, UK, Canada, Australia, and New Zealand). This generates a dataset of remote and hybrid work adoption whose scale, granularity, and high frequency extend well beyond what is possible to achieve with surveys. We use this to

zoom in on countries, occupations, cities, and firms and, in each case, document a high degree of heterogeneity in remote work adoption since the pandemic. Moreover, this heterogeneity is not simply a function of pre-pandemic conditions. For example, the incidence of remote and hybrid work across cities in 2019 explains relatively little of the cross-city increase in adoption since. We conjecture that the heterogeneity we document has its roots in myriad forces, including worker and firm preferences, competitive pressures in the labour market, and local norms. An important topic for future research, which our measures can help advance, will be to quantify the relative importance of these factors.

The data series in this paper are available through a companion website WFHmap.com which we will continue to update regularly going forward.

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, 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". Bookings 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.

Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei (2020). "Language Models are Few-Shot Learners". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., pp. 1877–1901.

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. 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].

Luktevich, Ben (2022). *BERT Language Model*. SearchEnterpriseAI. URL: https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model (visited on 07/08/2022).

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.

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 Nov 2025

	(1)	(2)	(3)	(4)
Country	Vacancies	Employers	Cities	Sources
United States	407,197,777	8,238,151	44,357	54,368
United Kingdom	99,792,669	2,158,983	6,490	20,227
Canada	25,872,609	1,485,311	10,402	14,225
Australia	14,013,029	494,564	980	16,642
New Zealand	3,931,167	161,920	426	7,670
All	550,807,251	12,538,929	62,655	113,132

Note: Reported counts pertain to the universe of online postings from January 2014 to November 2025, 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 includes 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.

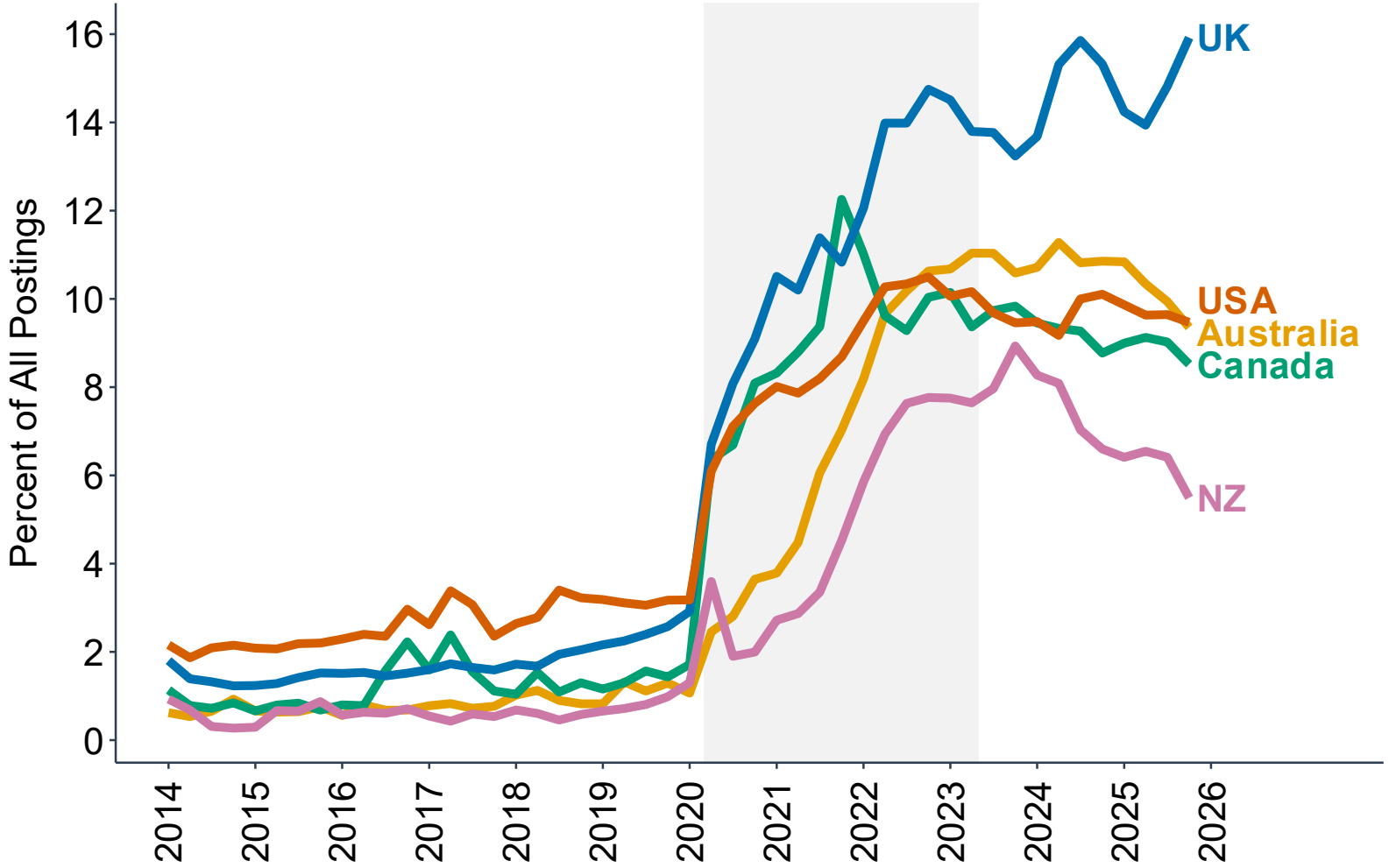
Note: The left column illustrates the role of attention weights in our 'Remote Work in job Openings' (RWO) large language model classifications of vacancy postings, where darker shadings pertain to higher weights. The right column illustrates the application of dictionary methods to the same text passages,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

Table 4: RWO Outperforms Other Classification Methods

	Audit Sample			Approximate Random Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
	Error Rate	Precision	F1 Score	Error Rate	Precision	F1 Score
All Zero	.28	.00	.00	.03	.00	.00
Dictionary	.16	.68	.74	.14	.15	.25
Dictionary w/ Negation	.12	.82	.78	.07	.28	.40
Logistic Regression	.11	.81	.81	.07	.26	.40
Logistic Regression w/ Negation	.08	.87	.85	.05	.36	.50
GPT-4o	.03	.95	.94	.02	.59	.73
Claude Sonnet 4	.04	.89	.93	.05	.38	.55
RWO (Generic English)	.03	.95	.95	.02	.66	.78
RWO (Baseline)	.02	.97	.97	.01	.75	.85

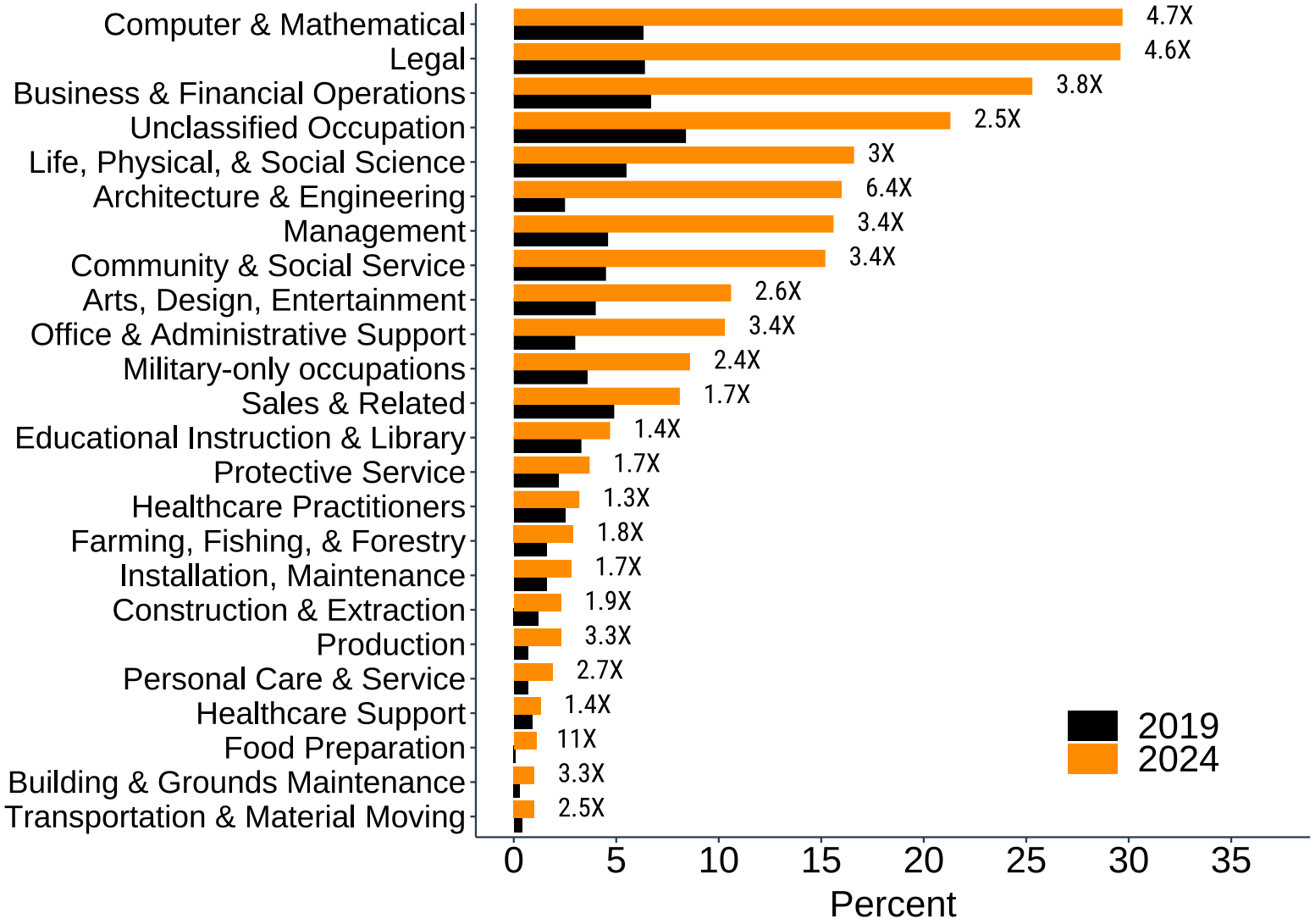
Note: This table reports classification performance metrics, which we calculate using a hold-out sample of human-classified text sequences. "Error Rate" is the overall rate of misclassifications (relative to humans). "Precision" is the ratio of true-positive classifications to the sum of true positives 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)-(3) uses a 40% random subset of our audit sample, and Columns (4)-(6) uses a sample that approximates a random sample of our full universe of postings. See Appendix B for details, including a description of each algorithm.

Figure 1: Vacancy Postings that Explicitly Offer Hybrid or Fully Remote Work Rose Sharply and Persistently



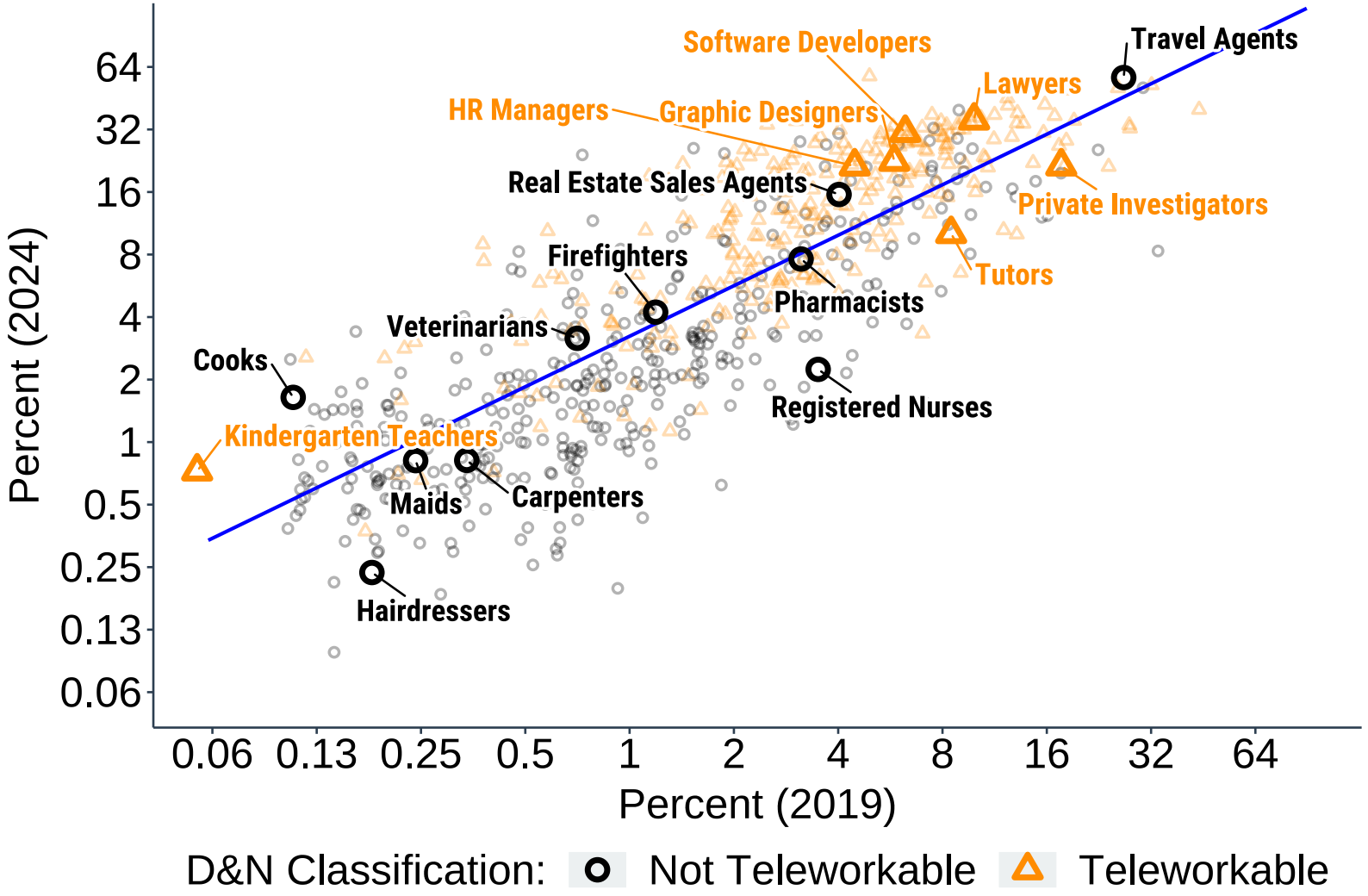
Note: This figure shows the percent of vacancy postings that say the job allows one or more remote workdays per week, encompassing both hybrid and fully-remote working arrangements). We compute these monthly, country-level shares as the weighted mean of the own-country occupation-level shares, with weights 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 2: Professional, Scientific and Computer-Related Occupations Have the Highest Shares of Postings that Offer Hybrid or Fully-Remote Work, U.S. Data



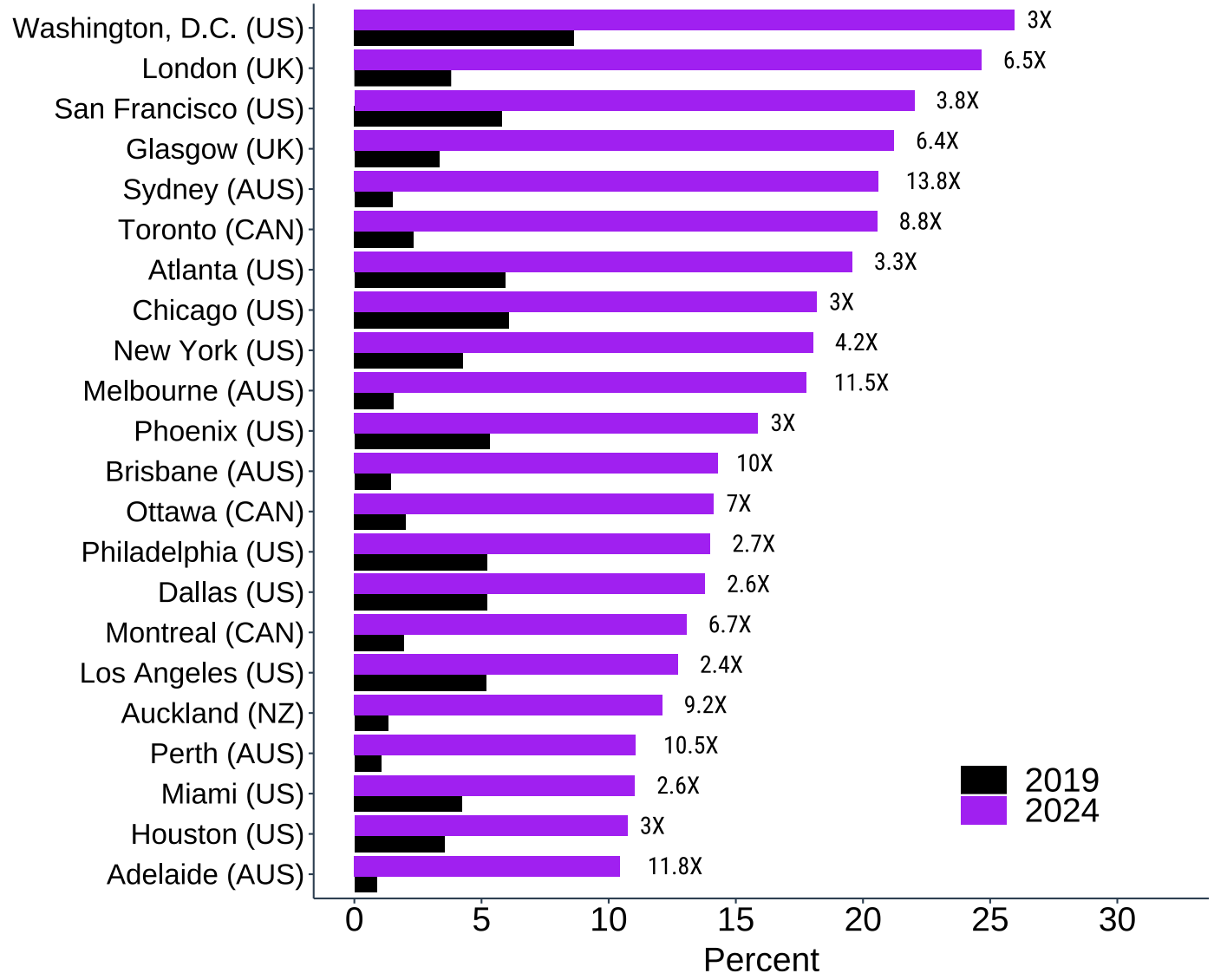
Note: Each bar reports the percent of vacancy postings that say the job allows one or more remote workdays per week in the indicated period and occupation group (two-digit SOC).

Figure 3: The Share of Vacancy Postings that Explicitly Offer Hybrid or Fully Remote Work Rose in Almost Every Occupation, U.S. Data



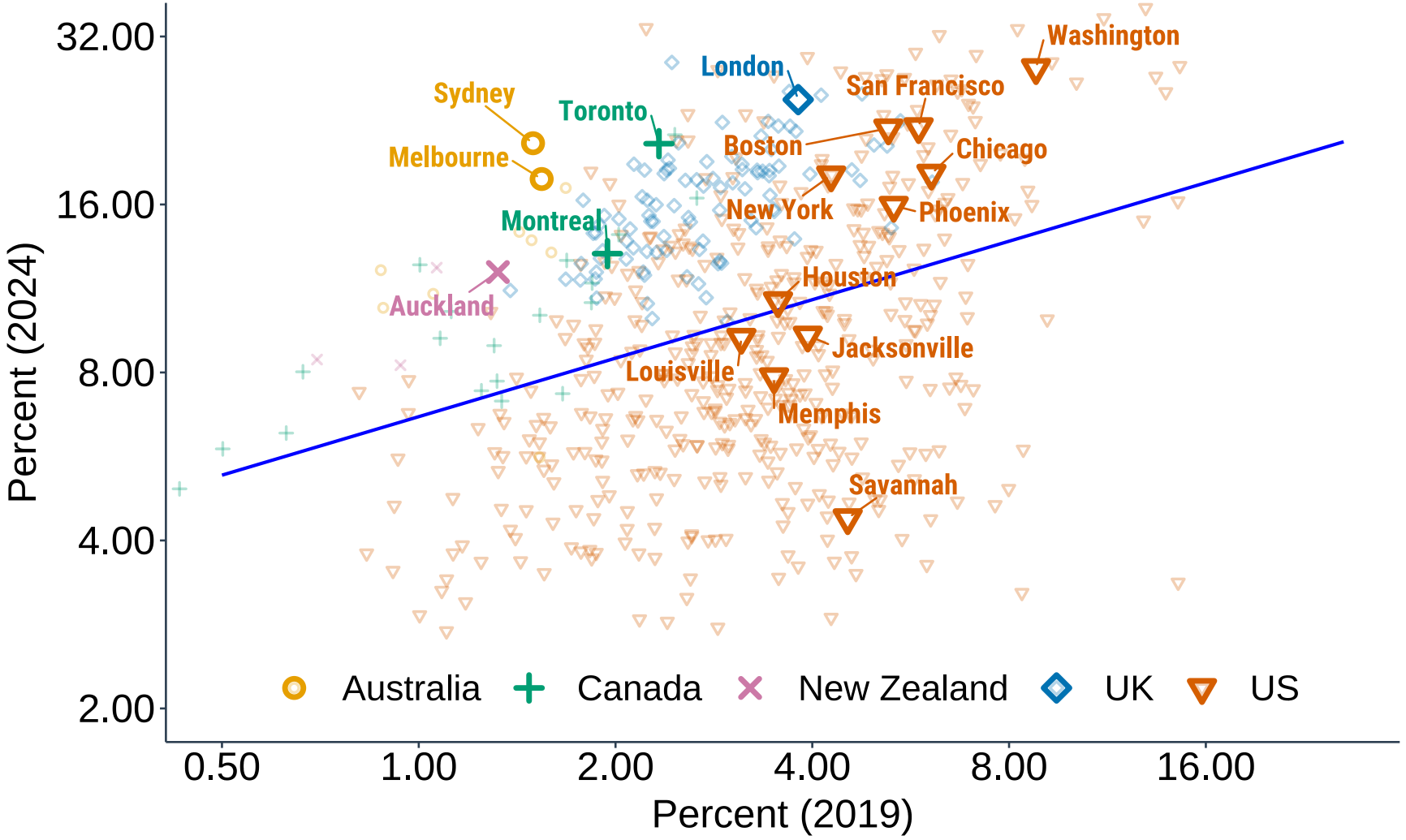
Note: This figure plots the percent of postings that say the job allows one or more remote workdays per week for 875 occupations in 2019 and 2024. We define occupations by ONET codes, dropping those with fewer than 250 posting in 2019 or 2024. The line shows the unweighted OLS fit: log(y) = 1.17 + 0.81 log(x), which has an R^2 value of 0.80. The color and shape denote whether Dingle & Neiman (2020) classify the occupation as feasible for fully remote working.

Figure 4: The Share of Vacancy Postings that Explicitly Offer Hybrid or Fully Remote Work Varies Widely across Major Cities



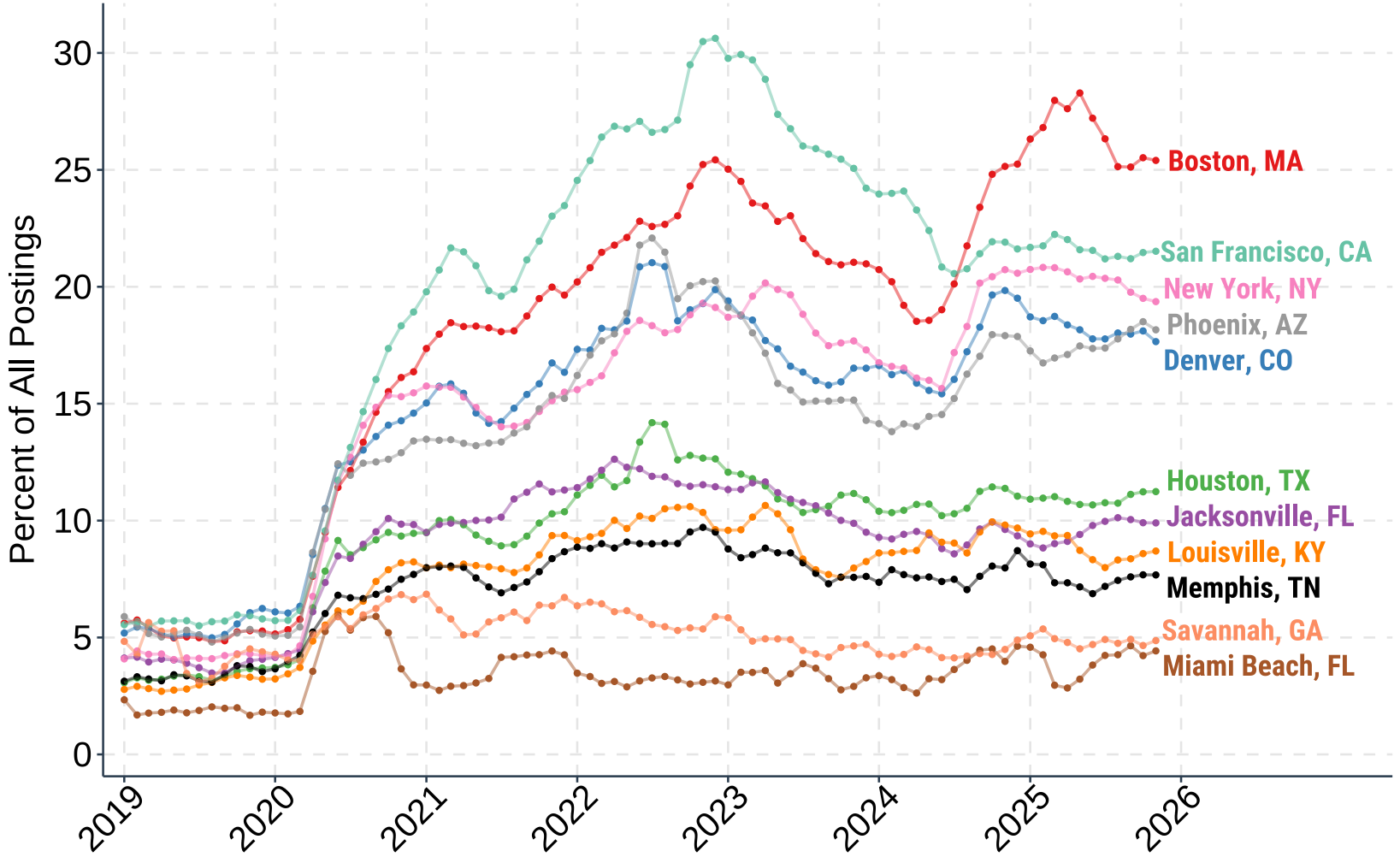
Note: Each bar reports the percent of vacancy postings that say the job allows one or more remote workdays per week in the indicated period and city. City refers to the location of the establishment or firm that is hiring.

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



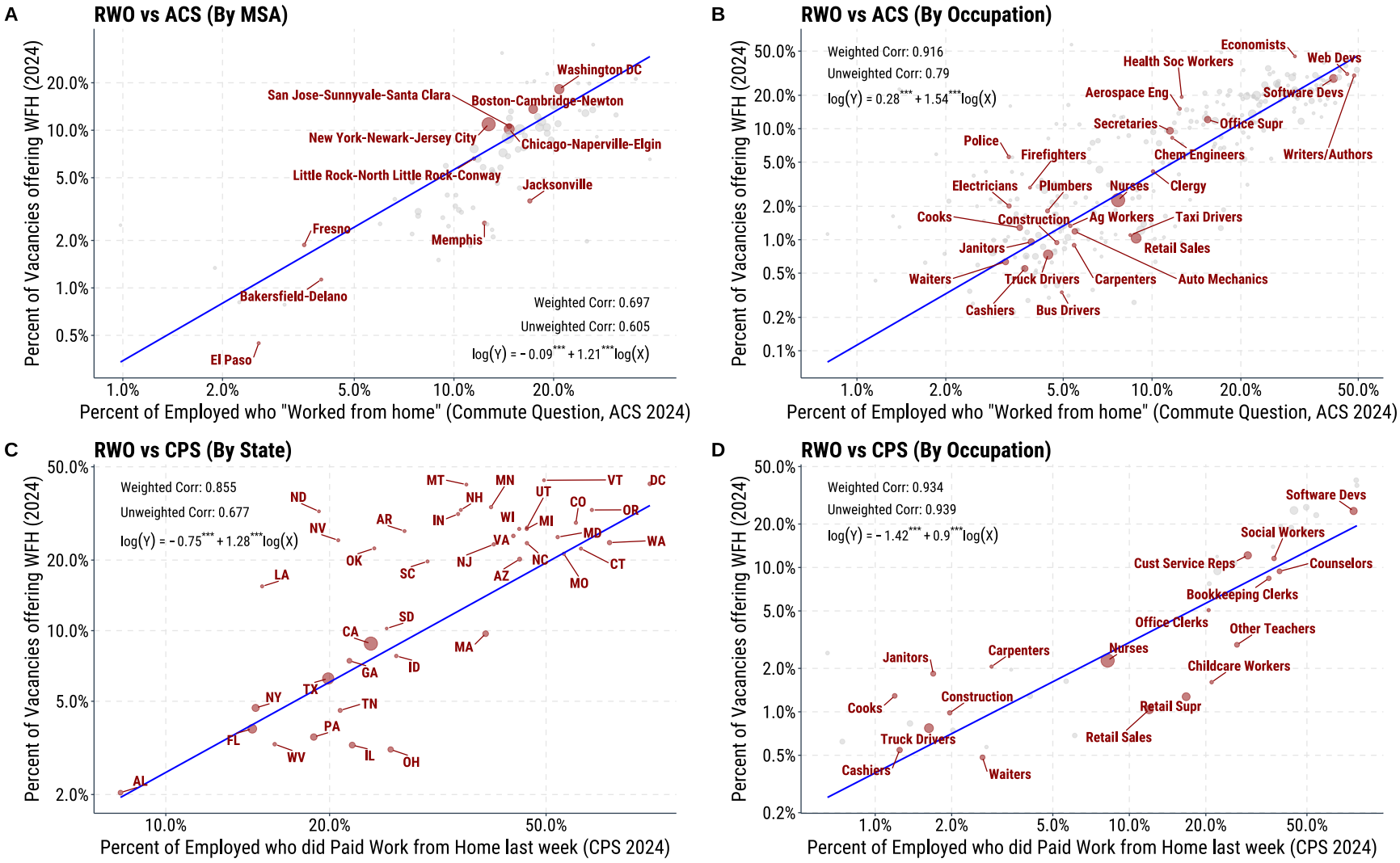
Note: This figure plots the city-level percent of postings that say the job allows one or more remote workdays per week in 2019 and 2024. "City" refers to the location of the establishment or firm that is hiring. The line shows the unweighted OLS fit: log(y) = 1.98 + 0.34 log(x), which has an R^2 value of 0.11.

Figure 6: Share of Postings Offering Hybrid or Fully Remote Work vary across US cities



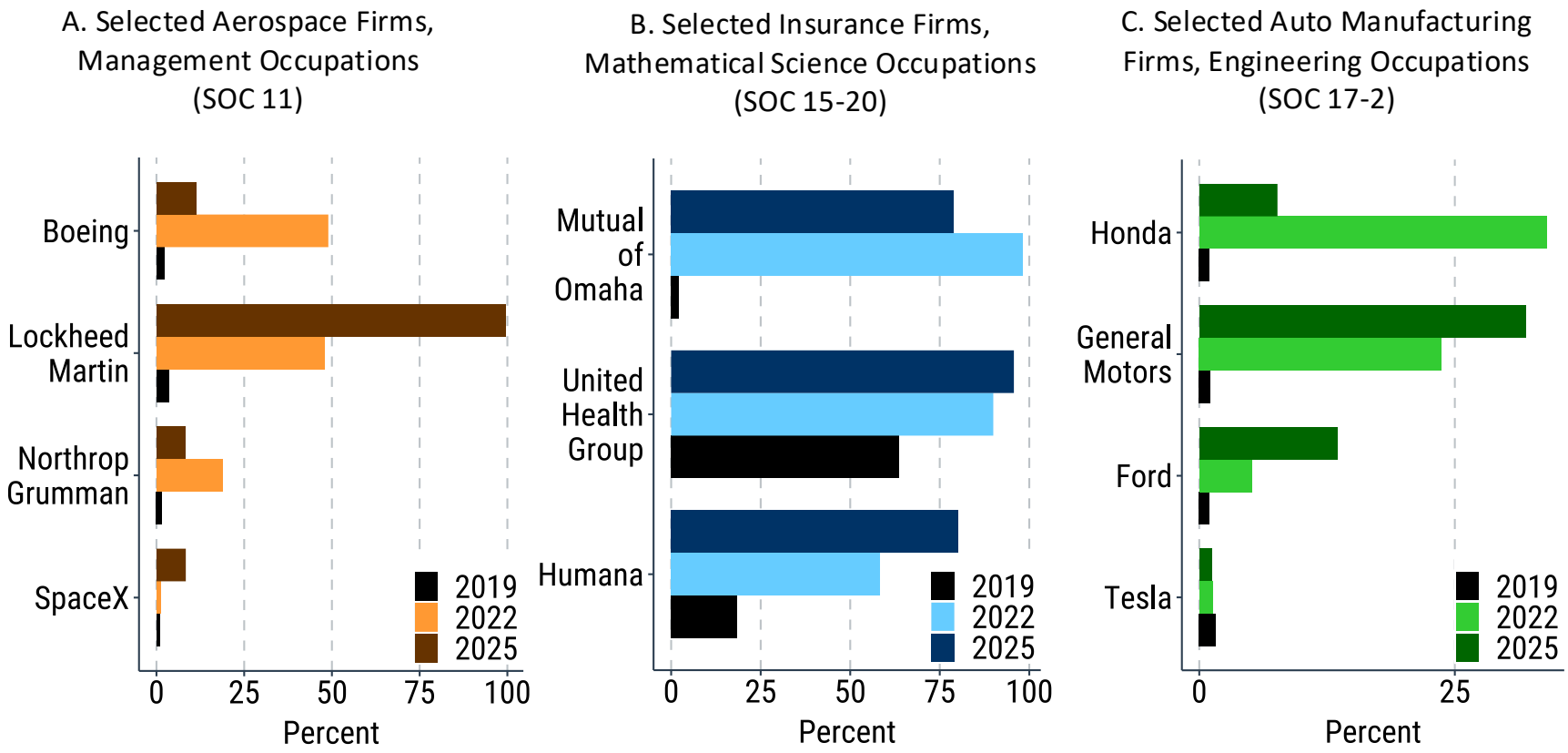
Note: We calculate the monthly share of all new job vacancy postings which explicitly advertise remote working arrangements (i.e. both hybrid and fully-remote), by selected cities. Prior to aggregation at the monthly level, we employ a jackknife filter to remove a small number of outlier days (see Appendix A: Data for further details). This figure shows the 3-month moving average. Cities chosen above are selected examples to illustrate the wide cross-city spread.

Figure 7: 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 8: 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.

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 manor.[29] 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 it 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.

[29] 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 predict 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 who 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.

B Supplementary Results

See figures below.

C Supplementary Information on Measurement

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.[30]

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 predicted 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.

2. **Tagging**: We search for the appearance of any of the keyword phrases from Table 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').

[30] 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.

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. (2022). 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 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

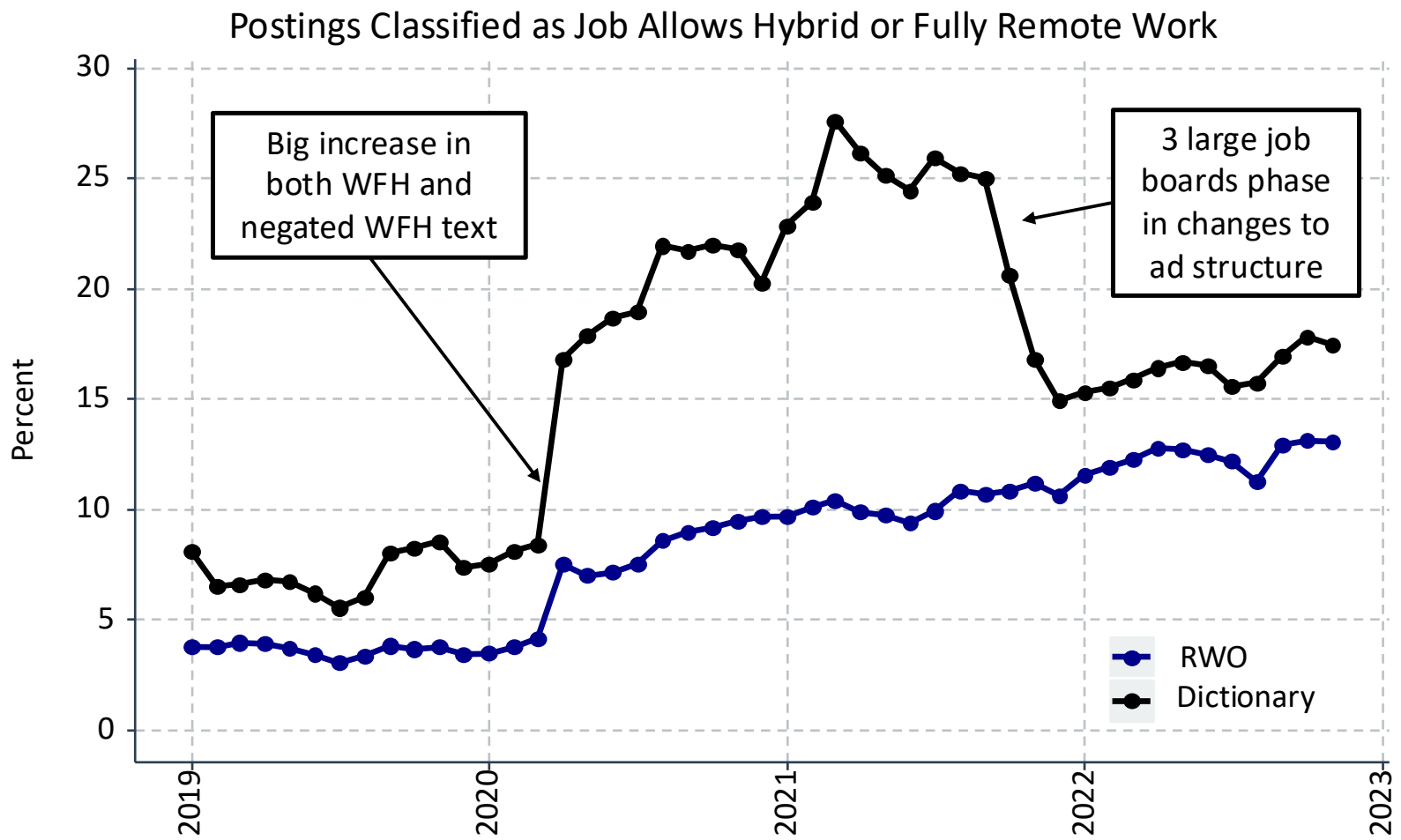
We follow an identical procedure to the one described for logistic regression but we further extend the document-term matrix with one extra column per keyword in the dictionary that indicates that the keyword was negated (according to our negation adjustment described before).

C.2.5 Zero-Shot Learning

The prompt for GPT-4o and Claude Sonnet 4 is equivalent and has the following structure:

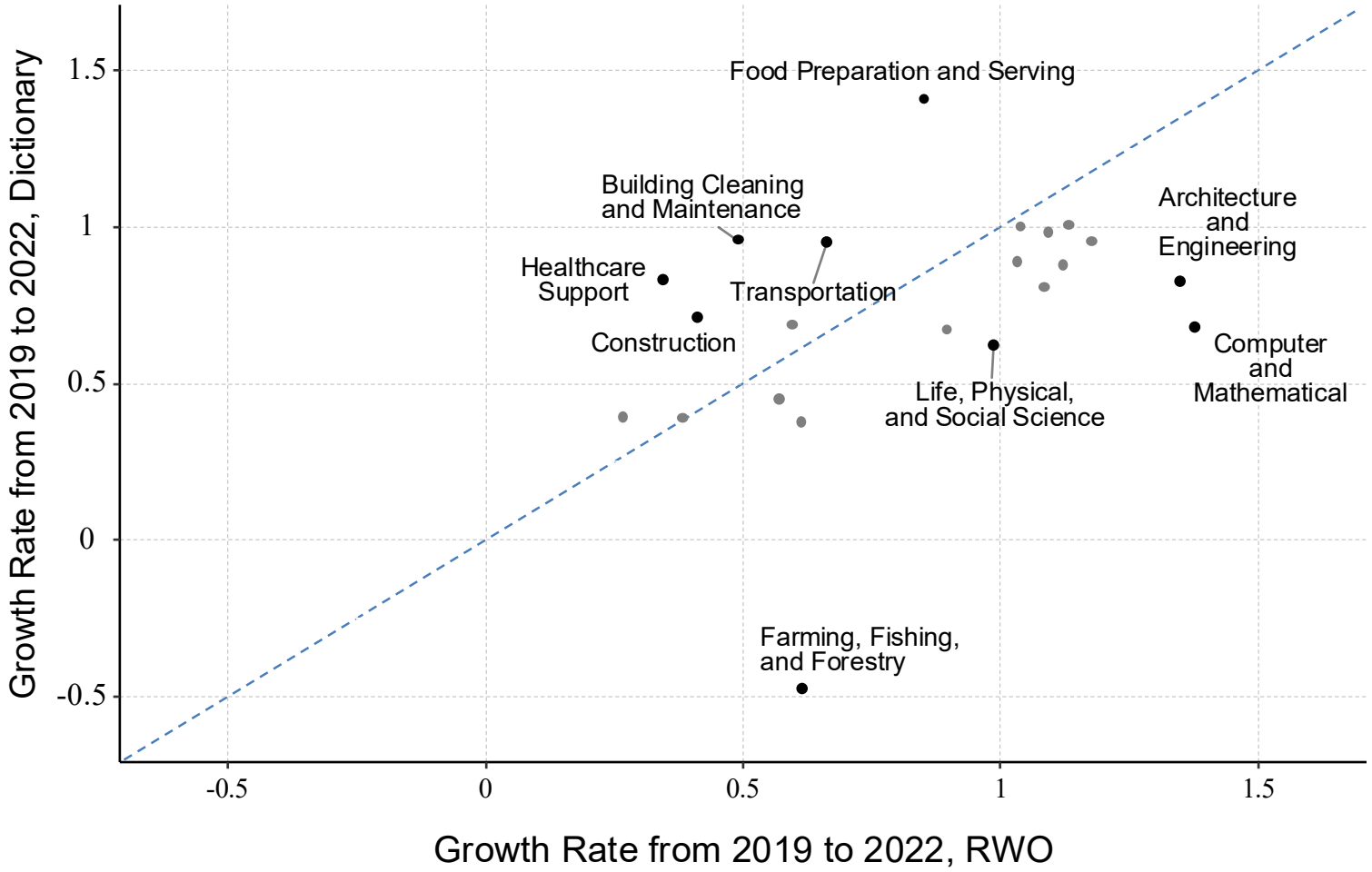
```
You are a data expert working for the Bureau of Labor Statistics.  You are
analysing the text of fragments of job postings and classifying them into
one of two categories:
0.  No remote work:  the text doesn't offer the possibility of working any
day of the week remotely.
1.  Remote:  the text mentions 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]
```

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



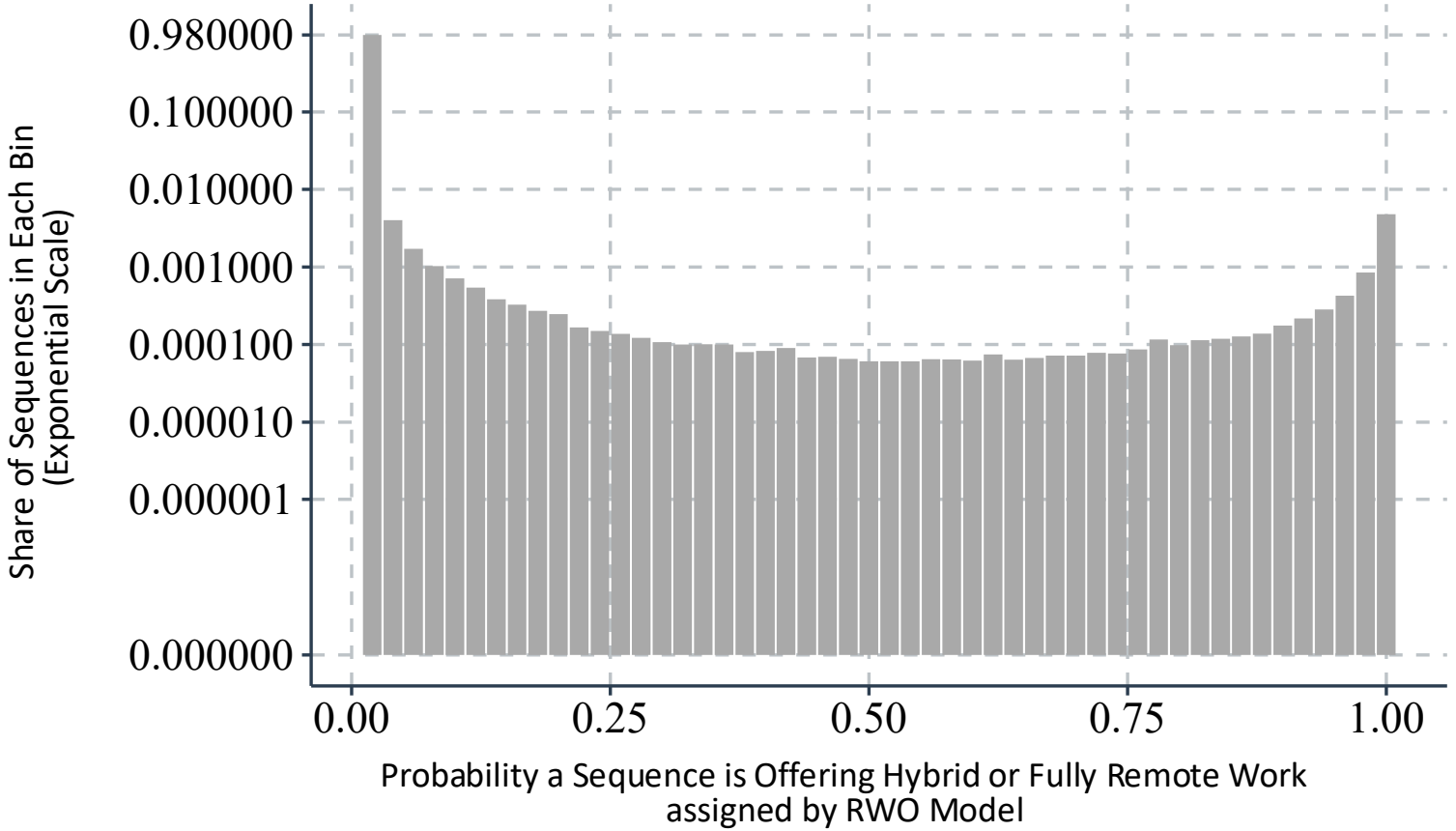
Note: This figure shows the percent of postings that say the job allows one or more remote workdays per week, as classified by our 'Remote Work in job Openings' (RWO) large language model (blue) and a dictionary-based approach (black) using the keywords in Adrjan et al. (2021). For both methods, we reweight 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	40,006,052	90.4
1	2,682,844	6.1
2	989,084	2.2
3	365,970	0.8
More than 3	201,523	0.5

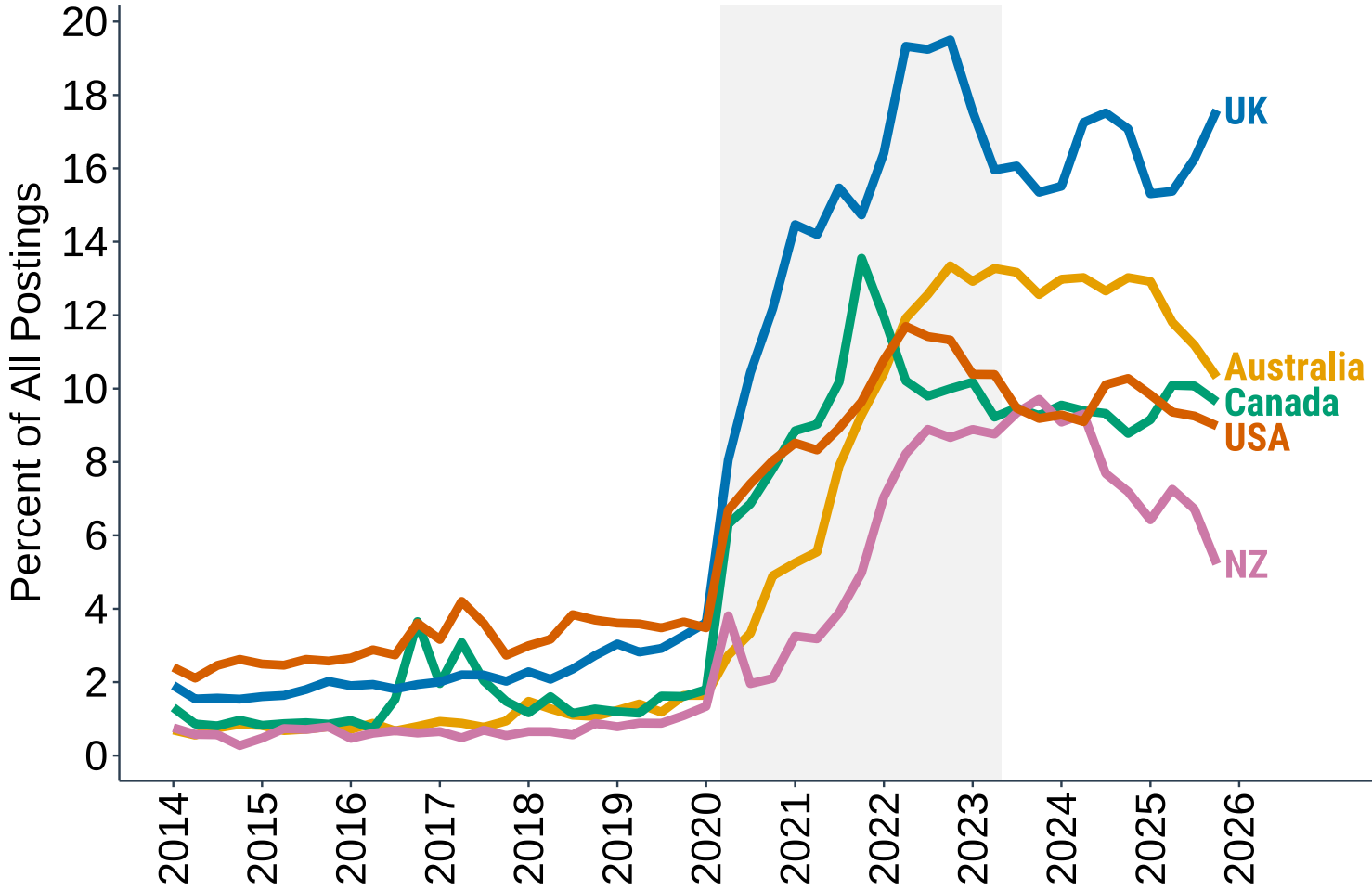
Note: This table tabulates how many text sequences in each US job posting from 2021 are classified as offering remote work (either hybrid or fully remote) according to our 'Remote Work in job Openings' (RWO) large language model. A typical job ad is split into six sequences. Most postings (90.42%) have no positive sequences. Of the remaining fraction, most have only one positive sequence.

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

	(1)	(2)	(3)
	Pretraining	Fine-Tuning	Full Sample Prediction
Computational setup	GCP Virtual Machine with 1 NVIDIA V100 GPU	GCP Virtual Machine with 1 NVIDIA V100 GPU	GCP Virtual Machine with 8 NVIDIA A100 GPUs
Total time (hours)	12	3	79
Job postings (per hour)	NA	NA	7,000,000
Cost per hour (USD)	$3	$3	$40
Total Cost (USD)	$36	$9	$3,147

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, Column (3) reports the cost for classifying just over 550 million job postings.

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 3 uses vacancy shares from the US to control for occupational composition across countries.

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 teleworking
remote work	remote	remotely
homeoffice	home office	home based
homebased		

Note: These are the keywords that appear in Table A.2 of Adrian 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.