# AI, firms and wages: Evidence from India*

Alexander Copestake[1], Ashley Pople[2], and Katherine Stapleton[3]

[1]International Monetary Fund

[2]University of Oxford

[3]World Bank

September 10, 2021

**Abstract**

We examine the impact of artificial intelligence (AI) on hiring and wages in service sector firms, using a novel dataset of vacancy posts from India's largest jobs website. We first document a rapid rise in demand for machine learning (ML) skills since 2016, particularly in the IT, finance and professional services industries. Vacancies requiring ML skills list substantially higher wages, but require more education and are highly concentrated both geographically and in the largest firms. Exploiting plausibly exogenous variation in exposure to advances in AI capabilities, we then examine the impacts of establishment demand for ML skills, as a proxy for AI adoption. We find that growth in the demand for ML skills has a direct negative impact on the total number of vacancies posted by incumbent firms. Drawing on rich data on wage offers, we further find that growth in ML demand reduces wage offers for all but the lowest-paid roles.

*Keywords:* artificial intelligence, labour markets, wages, development
*JEL Classification Codes:* J23, O33

*Email: acopestake@imf.org. For support and comments on this project, we are extremely grateful to Abi Adams-Prassl, Richard Baldwin, Stefan Dercon, Lukas Freund, Sanjay Jain, Simon Quinn, Ferdinand Rauch, Chris Woodruff, as well as various seminar participants. We are indebted to Sanjay Jain, Sanjeev Bikhchandani, Pawan Goyal, Abhishek Shyngle and Shweta Bajad for help with the vacancy data, and to Rob Seamans for sharing the measure from Felten et al. (2019). Max Marczinek provided outstanding research assistance. We gratefully acknowledge financial support from Oxford Centre for the Study of African Economies (CSAE). The views expressed herein are those of the authors and should not be attributed to any of the institutions with which the authors are affiliated.

# 1 Introduction

Advances in artificial intelligence (AI), driven by progress in the sub-field of machine learning (ML), have spurred an intense debate about the impact of AI on jobs.[1] Yet, despite widespread discussion, detailed empirical evidence measuring this impact remains limited. This is particularly the case for middle- and low-income countries, where little is known to date about the extent of deployment of AI and what it is being used for, let alone how it is affecting labour markets. For countries pursuing a services-led development model, this question has broad ramifications: many of the services industries that have driven growth and job creation, such as Business Process Outsourcing (BPO), are now highly susceptible to ML-based automation. For instance, in India – the archetype of the services-led development path – the IT-BPO sector currently employs around 4 million people and contributes 8% of India's GDP (SESEI 2019).

The theoretical impact of AI on jobs is ambiguous. Advances in ML have reduced the cost, or improved the quality, of the task of 'prediction', which is prevalent in many occupations (Agrawal et al. 2018).[2] While these trends initially suggest displacement of labour, improvements in the task of prediction could also expand labour demand by reducing overall costs of production or increasing quality and hence raising productivity.[3] In addition, AI could complement human labour, create entirely new tasks or incentivise changes in organisational structure; indeed, there is growing evidence that AI is a general-purpose technology (GPT), an 'invention of a method of invention' (Brynjolfsson et al. 2017, Cockburn et al. 2018, Klinger et al. 2018, Goldfarb et al. 2020, Agrawal et al. 2021).[4] Emerging economies like India could benefit from new global AI value chains, capitalising on their abundance of engineering talent, existing expertise in IT outsourcing, and further declines in communications costs (Baldwin & Forslid 2020*a*). Indeed, revenues in India's BPO sector nearly tripled over the past ten years (NASSCOM 2018).

In this paper, we investigate the labour market impact of AI in the services sector in India

[1] To fix definitions, we consider AI 'the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages' (Oxford English Dictionary 2020). ML, the sub-field responsible for many of the recent commercial applications of AI, comprises 'the statistical techniques that enable computers and algorithms to learn, predict and perform tasks from large amounts of data without being explicitly programmed' (Acemoglu & Restrepo 2019). We henceforth use 'AI' as an umbrella term encompassing ML.

[2] For example, a back office employee of a multinational bank takes the input of scrawled handwriting on a mortgage application form, then generates the correctly spelled name of the applicant as predicted output.

[3] Some early research has therefore modelled ML in a comparable way to other forms of automation, such as industrial robots (e.g. Webb (2020) and Acemoglu et al. (2020)). These papers build on the canonical framework of Acemoglu & Restrepo (2018) in which task structure determines adoption. Beyond the boundary of the adopting firm, AI could also have broader indirect effects, as workers reallocate across occupations – as explored in detail by Humlum (2019) in the case of robots. Here we focus on direct within-firm effects.

[4] Specifically, GPTs (1) are widely used across sectors, (2) have inherent potential for technical improvement, and (3) spawn further innovation in application sectors (Bresnahan & Trajtenberg 1995).

using a novel dataset of vacancy posts from the country's largest jobs website, estimated to have around a 60 percent market share of Indian online job postings. The dataset has wide coverage of vacancy posts for new hires of white-collar service sector workers. Following Acemoglu et al. (2020) and Stapleton & O'Kane (2020), we gauge firm-level AI adoption using demand for machine learning skills, as observed in the text of posted job descriptions. Our data also include firm identifiers, locations, salary offers and skill, experience and education requirements. Moreover, all job postings require information on wage offers, which is largely incomplete in other scraped job postings data found in other contexts, such as Burning Glass Technologies. As a result, the data provide a granular insight into the Indian labour market, allowing us to explore the impact of AI on other detailed proxies for labour demand.

Using the vacancy data, we first document several trends in AI adoption in India's service sector. We see a rapid take-off in 'AI demand' (shorthand for the demand for AI-related skills in vacancy posts) after 2016, particularly in the IT, finance and professional services industries. AI demand increased from 0.37% of all job vacancies in 2015 to 1.03% in 2019, coinciding with a rapid increase in demand for 'deep learning' skills, along with 'natural language processing' to a lesser extent. AI roles tend to require substantially more education, particularly graduate degrees, while also paying significantly more. Even after controlling for detailed region, industry, firm, occupation and role fixed effects, posts demanding AI skills still pay a 13 to 17% salary premium. AI roles are heavily concentrated in a few key technology clusters – particularly Bangalore, Mumbai, Hyderabad, Pune, Chennai and Delhi – and in the largest firms. Consistent with this spatial clustering, we find evidence of local diffusion: after the first firm in a given industry and region adopts AI, other firms in the same industry and region are on average more likely to start demanding AI skills, even after taking into account industry and region trends.

We next consider the relationship between AI demand and total labour demand. As noted above, AI adoption could lead to many effects across the economy. We aim to identify one narrow subset of these impacts empirically: the net direct effect of adopting AI on hiring within pre-existing firm-city pairs (hereafter 'establishments'). Using a long difference specification between 2010-12 and 2017-19, we investigate the effects of growth in the demand for AI skills on the growth of non-AI job postings and wage offers at the establishment level. To isolate causation, we exploit establishment-level variation in exposure to supply-side advances in AI capabilities developed outside of India, as reflected in the exposure measure of Webb (2020). Specifically, this measure captures the degree of overlap between occupations' tasks and the tasks which patented AI technologies are designed to perform. We aggregate this measure to the establishment level using establishment occupation vacancy shares at baseline, then use this variation as an instrument for AI demand.[5]

We first examine the first stage and find that firms more exposed to AI *ex ante* see a relative increase in their demand for AI skills in online vacancy posts. Turning to the second stage, we find that growth in AI demand has a significant negative effect on growth in non-AI and total postings by establishments. A 1% increase in the AI vacancy growth rate results in a 3.61 percentage point decrease in establishment non-AI vacancy growth between 2010-12 and 2017-19, controlling for region, firm size and industry fixed effects. Growth in total establishment vacancies (AI plus non-AI) falls by a similar 3.57 percentage points, reflecting that the increase within the small set of AI posts is far outweighed by the displacement effect in the larger set of non-AI vacancies.

How does this displacement affect wage offers for new hires? We find that a 1% higher growth rate in AI vacancies reduces the growth rate of non-AI median wage offers by 2.6 percentage points between 2010-12 and 2017-19, instrumenting with AI exposure and controlling for region, firm size and industry fixed effects. The negative effect persists even when controlling for changes in the education and experience profile of jobs, although the average level of experience demanded in the non-AI hires also declines slightly. This suggests that the decrease in the wage offer growth rate is not solely attributable to AI-adopting firms demanding lower-skilled workers. These negative effects appear across the wage distribution, with strong statistically significant effects above the 20th percentile and imprecisely estimated effects at the lowest decile.

These results are robust to various alternative econometric specifications, including using mean wages and weighting by establishment size (excluding outliers). The results also receive some support from alternative measures of AI exposure.[6] Finally, traditional sources of labour market data, such as the National Sample Survey and Periodic Labour Force Survey, also reveal consistent patterns at the industry-district level: there is a strong negative relationship between AI exposure and wage offer growth.

This paper makes several contributions to the literature. Firstly, we offer detailed new evidence on the demand for AI skills in India's service sector firms using a novel dataset. Building on studies using vacancy postings to assess AI adoption in the US and UK (Alekseeva et al. 2020, Acemoglu et al. 2020, Stapleton & O'Kane 2020), we find that growth in the demand for AI

[5]To isolate the impact of AI *usage*, rather than AI *production*, we exclude AI-producing sectors from our analysis – specifically IT and education, which are responsible for the vast majority of AI patents (Klinger et al. 2020).

[6]We use the Webb version as our main measure, because it most closely captures supply-side advances in the deployable capabilities of AI. We also consider the measures of Felten et al. (2018) and Mani et al. (2020), the latter of which applies the methodology of Brynjolfsson & Mitchell (2017) to India. Our first stage and wage results are very similar with the Felten et al. measure. In contrast, the Mani et al. measure may be capturing a different phenomenon, and does not predict firm demand for AI skills. See Section 6 for a detailed discussion.

skills in India has been similarly rapid, with a take-off around 2016 and broad adoption across industries. Such similarity across countries and income levels adds further support to existing evidence that AI is a GPT (Goldfarb et al. 2020). Our findings for India are also consistent with Alekseeva et al. (2020) in showing that roles requiring AI skills now provide a substantial wage premium, with our estimate of 13-17% just above their estimate of 11%. In addition, our firm identifiers enable us to extend the literature by documenting the degree of concentration of AI demand in large firms, showing that this has been high and rising over time for India.

Second, we offer an early attempt to evaluate the effects of establishment-level AI deployment, proxied by the demand for ML skills, on the establishment's labour demand and wages. Similar to evidence from the US (Acemoglu et al. 2020), but unlike evidence from the UK (Stapleton & O'Kane 2020), we find a substantial negative relationship between AI demand and non-AI labour demand at the establishment level. Our IV strategy also allows us to isolate the direct effects of AI adoption, by exploiting variation from those establishments induced to demand AI skills as a result of their exposure to AI technological advances. This contrasts with other work studying solely the effects of AI exposure, such as Webb (2020) and Acemoglu et al. (2020).[7] Additionally, unlike in the existing studies using Burning Glass Technologies (BGT) data, our vacancy data also includes comprehensive information on wage offers, allowing us to study the effects of wage offers for new hires. Together, these innovations allow us to identify new insights on the significant negative effects of AI demand on wage offers.

Third, we aim to bridge a significant evidence gap on AI adoption and its labour market effects in low- and middle-income countries. Discussions of AI and development macroeconomics have been largely theoretical to date. For instance, Baldwin (2019) and Baldwin & Forslid (2020*b*) have conjectured that machine learning, along with online platforms and software robots, could benefit developing countries by increasing offshoring of services. Korinek & Stiglitz (2021) take an alternative view that developing countries will be negatively affected, because AI devalues their comparative advantages in abundant labour and natural resources. On the one hand, our finding that the rapid deployment of AI in India's services sector has had labour market effects at least as negative as those observed in high-income countries appears more in line with Korinek & Stiglitz (2021). On the other hand, our negative findings only concern within-firm effects for incumbent firms; we may observe offsetting effects through other channels, such as firm creation. Indeed, our focus on 'AI-using' industries means we exclude positive employment effects of 'AI production', particularly in the IT industry.

[7] Our first stage mirrors the tests of Proposition 1 in Acemoglu et al. (2020), while their tests of Proposition 2 are analogous to the reduced form of our estimates of non-AI vacancy growth.

Finally, our paper adds to a growing literature which uses online vacancy postings to investigate labour market effects more broadly (e.g. Deming & Kahn 2018, Stapleton & O'Kane 2020, Adams et al. 2020, Javorcik et al. 2020). We contribute through a large new dataset of job posts in India, stretching back a decade to 2010. Here, we also build upon Chiplunkar et al. (2020) who use five months of 2020 data from India's second-largest job portal to study the impact of the COVID-19 pandemic.

The rest of this paper proceeds as follows. Section 2 introduces the data. Section 3 then presents detailed descriptives on AI demand in the Indian white-collar services sector. Section 4 outlines the main empirical approach, Section 5 presents our results on the relationships between AI demand and broader job postings outcomes, and Section 6 addresses their robustness. Section 7 concludes. Appendix A contains additional figures and tables, and Appendix B provides a detailed description of the vacancy dataset.

# 2 Data

## 2.1 Vacancy data

Our primary dataset is the online job vacancy data posted on India's largest national job board platform between 2010 and 2019.[8] The online job board serves primarily as an advertising platform for firms to post vacancies, with subsequent recruitment and hiring processes taking place directly with firms. The platform estimates that they had approximately 60 percent of the market share of Indian online job vacancies in 2020. They shared a randomly-selected 80 percent of all posts over the period 2010 to 2019. We focus on the services sector, for which the data is most representative of overall job vacancies, dropping posts from the manufacturing and agriculture sectors where the coverage is thin.[9] Our primary dataset hence includes text data from around 15.5 million service sector job postings, equating to roughly 1.5 million per year, but skewed towards the end of the sample. In our main analysis, we further restrict our sample to a panel of firms that have posted at least one job between 2010 and 2012 and again in the 2017 to 2019 period. This gives a final sample of approximately 2 million posts.

As illustrated in Appendix Figure A.1, there is relatively broad coverage of vacancy posts by geography in India, with posts concentrated in urban centres. Over 150,000 unique firms posted at least one vacancy over the ten-year period, with an average of 80 posts per firm. We provide

[8]The company requested to remain undisclosed.

[9]This focus aligns with our core question, namely how the adoption of new AI technologies might affect the prospects for services-led growth models.

a detailed discussion of the data in Appendix B, including its representativeness in Appendix B.2, where we benchmark the vacancy data relative to nationally-representative labour surveys (the 2011-12 National Sample Survey and 2017-18 Periodic Labour Force Survey) and firm-level data from Prowess. Our dataset is substantially larger than these alternatives. For instance, in 2018 we have more than two million white-collar services vacancies from 40,000 unique firms, compared to 12,000 workers directly surveyed in the 2017-18 Periodic Labour Force Survey and 2,000 firms recorded in Prowess. Full observation totals by sector and dataset are shown in Table B.1.

Our vacancy data has several advantages over the administrative datasets. First, the representative sample surveys only took place in 2011-12 and 2017-18, so provide no information on short-term fluctuations or more recent developments in the Indian services sector. Prowess, while useful for studying the largest firms, only contains a limited selection, and did not yet have good coverage for recent years at the time of writing. As illustrated in shown in Appendix Figure B.1, our vacancy dataset has roughly 30 times the number of firms compared to Prowess. Neither Prowess nor the labour surveys offer a clear window on AI exposure or adoption. In contrast, we can directly observe the demand for AI skills in online job descriptions.

When submitting a job vacancy on the platform, firms are required to upload information into a standardised template. Hence, all posts include information on the job title, industry, role category, location, skills required, salary and experience ranges and educational requirements. The job postings also include an open text section for the job description. We manually map industries and occupations into the National Industrial Classification (NIC) at the two-digit level and National Classification of Occupations (NCO) at the four-digit level, covering 99% of all vacancies. We also harmonise city names and added geolocations, separating out oversees job postings. Using the geolocations, we matched cities to districts, using the 2011 census. Firm names were removed for anonymity and replaced with a consistent panel identifier.[10] Appendix B provides further details on our approach to cleaning the vacancy data and benchmarking the data against administrative data.

## 2.2 Measuring AI demand

Despite the prominence of the topic of AI in popular discussion, firm-level data on AI adoption remains scarce (e.g. see more discussion in Seamans & Raj (2018)). In the absence of data on the adoption of specific technologies, a growing body of work has started using technology-related

[10] We also focus on full-time jobs and drop the small number of part-time and non-permanent positions from our sample.

human capital to proxy for technology adoption. For example, Rock (2019) and Benzell et al. (2019) use LinkedIn profiles to construct firm-level measures of engineering and IT talent, whereas Harrigan et al. (2016) use the firm-level employment share of 'technology workers' in French matched worker-firm data as a measure of technology adoption.

Human capital is one of the key inputs for deploying an AI system. It is well recognised that one of the primary obstacles to widespread adoption of AI is the available labour supply, with top-tier scientists earning extremely high salaries and being bought out of academic positions. It would be expected that firms wishing to implement an AI driven automation project would, at least to some degree, need to hire individuals with AI skills or experience. Alternative options would be to rely on external consultants, contract out the process to a third party software provider or retrain existing staff to develop AI skills.

There are a number of reasons to believe that the dominant channel is external hiring. Work by McKinsey Global Institute (2019) surveying around 2000 companies globally found that the primary method for sourcing AI talent and capabilities was to hire externally and that the majority of companies built their AI capabilities in house, as opposed to buying or licensing capabilities from large technology companies. Additionally, even if firms were to subcontract AI development, it would be expected that they would still require at least some related human capital in-house to oversee and manage the process. While we cannot rule out these other channels, we hence assume that AI skills demand and actual AI deployment within a firm will be at least moderately correlated and follow this emerging literature in using the demand for AI skills as a proxy for the extent of AI adoption.

Online job vacancy data lends itself well to the measurement of demand for very specific technology-related human capital owing to its detailed textual data on the skills demanded for roles. We therefore measure firm demand for AI skills through job vacancies as a proxy for AI adoption. To measure firm demand for AI skills, we classify job postings based on the text in the job description or skills requirements. Our main classification is the 'narrow' measure employed by Acemoglu et al. (2020), which categorises a post as an AI vacancy if it includes any word from a list of specific AI terms.[11] By using this narrow measure of AI skills, we reduce measurement error, although our estimates of demand for AI skills are likely to be a lower bound

[11]Specifically, a post is categorised as AI-related if any of the following terms appear in either the 'job description' or 'skills required' fields: Machine Learning, Computer Vision, Machine Vision, Deep Learning, Virtual Agents, Image Recognition, Natural Language Processing, Speech Recognition, Pattern Recognition, Object Recognition, Neural Networks, AI ChatBot, Supervised Learning, Text Mining, Support Vector Machines, Unsupervised Learning, Image Processing, Mahout, Recommender Systems, Support Vector Machines (SVM), Random Forests, Latent Semantic Analysis, Sentiment Analysis / Opinion Mining, Latent Dirichlet Allocation, Predictive Models, Kernel Methods, Keras, Gradient boosting, OpenCV, Xgboost, Libsvm, Word2Vec, Chatbot, Machine Translation and Sentiment Classification.

of the true level of adoption.

# 3 Descriptives

In this section, we present five key descriptive findings about the demand for AI skills in the Indian white-collar services sector, using the vacancy data. Additional figures and tables are provided in the Appendix A.

***1. AI demand increased rapidly after 2016, particularly in the the IT, finance, education and professional services sectors.***

AI demand increased rapidly after 2016, rising from 0.37% of all job vacancies in 2015 to 1.03% in 2019. Figure 3.1 Panel (a) shows the share of posts that are tagged AI posts and the disaggregation of AI terms to examine which particular skills are most in demand. Panel (b) shows the share of AI posts over time and within the top five industries by AI share. Overall we see steady growth from our base year of 2010, which accelerates after 2016, especially in financial services and IT & software. This growth is driven largely by demand for general 'machine learning', with the sub-field 'deep learning' rising rapidly from relative obscurity to being the second most sought-after AI skill from 2017 onwards. Tellingly, the rapid take-off in the demand for AI skills after 2016 almost exactly matches patterns found in the US and UK by Acemoglu et al. (2020) and Stapleton & O'Kane (2020), respectively.[12]

The breakdown of demand for AI skills across sectors is also striking, suggesting a wider diffusion of AI beyond AI-producing sectors. AI demand grew steadily in the IT sector since 2011, whereas AI demand in the financial sector started from a low base and grew by ten-fold between 2016 and 2018. In contrast, the business process outsourcing and call centre sector saw a small boom in AI demand in earlier years, before petering off. In 2015 and 2016, this sector had the second-highest share of AI demand, corresponding to a rise in the demand for 'speech recognition'. The subsequent decline in the sector's AI share is therefore surprising.[13]

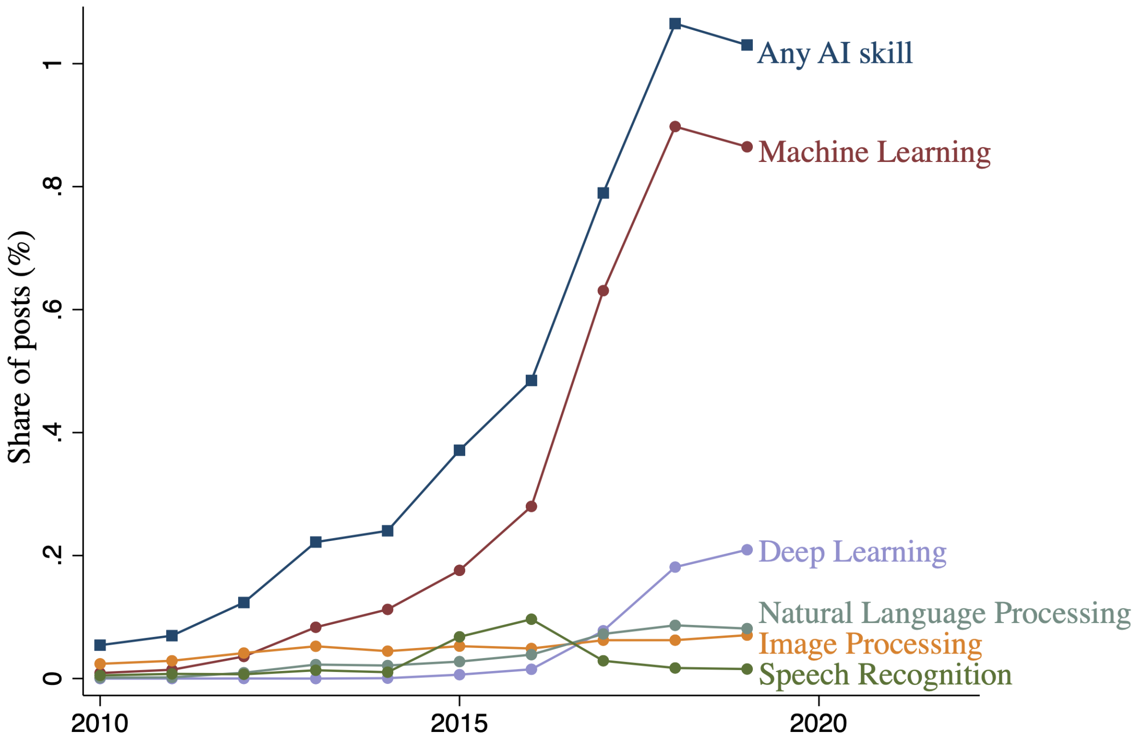
***2. AI roles require more education, but offer substantially higher wages than other white-collar services jobs.***

[12]For discussion of possible causes of the rapid acceleration, such as the open-source release of Google's TensorFlow software library, see Stapleton & O'Kane (2020).

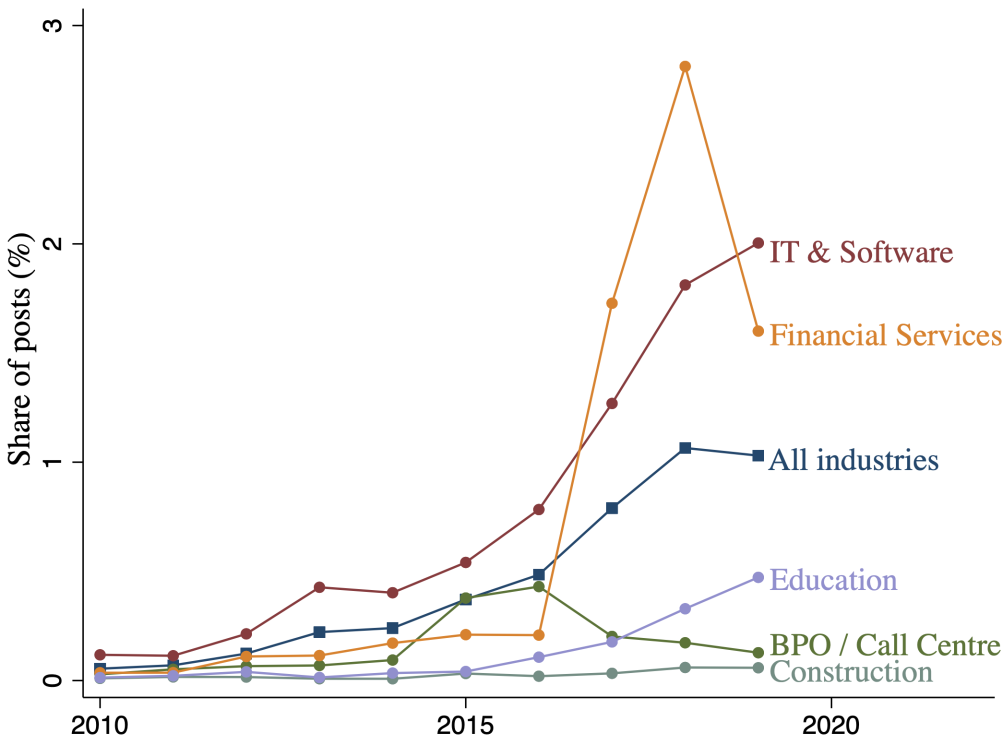
[13]Looking at the components of the share measure, we see that the absolute number of AI posts in the sector fell substantially in 2017 (see Figure A.3 in the Appendix).

Figure 3.1: Trends in AI demand

(a) Most demanded AI skills

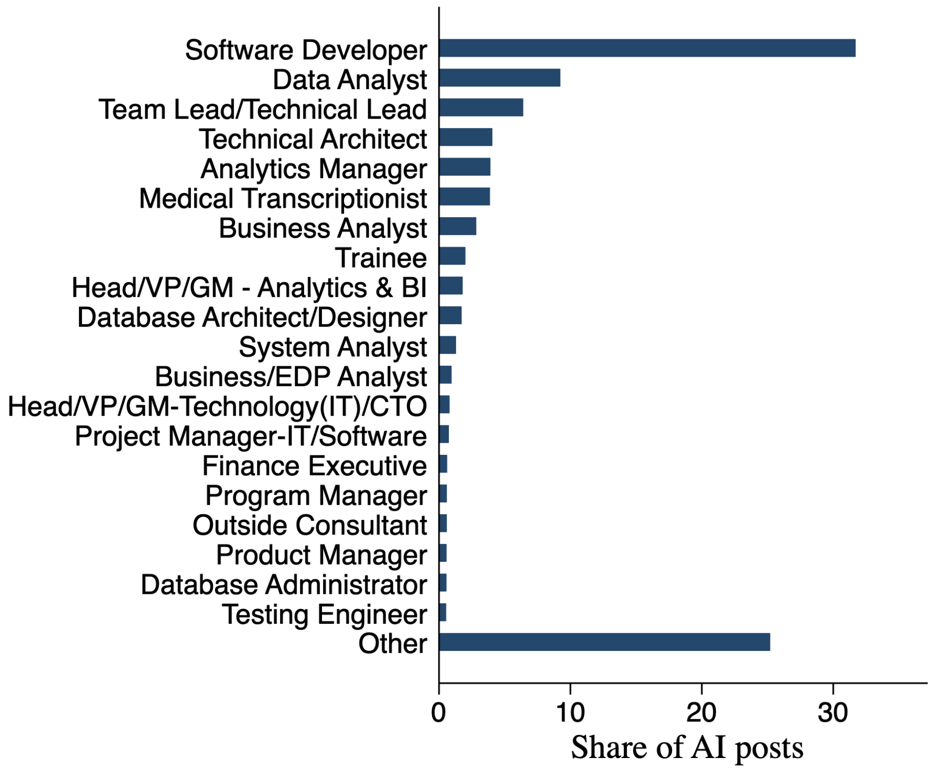



(b) AI share of posts, by industry




*Notes:* Panel (a) shows the share of all vacancies that specify particular AI skills, for the top five most demanded skills. Panel (b) shows the share of vacancies that are AI vacancies, both for all industries together and within each of the top five industries by AI share.

Figure 3.2: Top 20 roles demanding AI skills, 2010-2019




*Notes:* We rank the top roles demanding AI skills, by their share of AI posts. All other roles hiring AI skills are grouped in the 'Other' category.

What are these AI roles, and how do they compare to the rest of the vacancies advertised? By far the most common AI role title is 'Software Developer', followed by other technical roles such as 'Data Analyst', 'Technical Lead' and 'Technical Architect' (Figure 3.2). AI skills are also required in technical management roles, with titles as 'Analytics Manager', 'VP - Analytics & BI', and 'Project Manager-IT/Software' also appearing in the top 20 AI-related roles. Yet there is also a long tail of more generalist roles, including 'Business Analyst', 'Trainee', 'Program Manager' and 'Product Manager'. The size of the 'Other' category grouping all other vacancy posts (25%) indicates how widespread the hiring of AI skills is across multiple job titles, albeit each with a small share of overall posts.

Moreover, we see that AI-hiring firms are seeking candidates who are slightly more experienced and substantially more educated than average – and for that they are willing to pay a substantial salary premium (Figure 3.3). AI vacancies are almost twice as likely as non-AI vacancies to require a master's degree, and more than seven times more likely to require a doctorate. They post a median salary of ₹250,000 (approximately US$3,333, without adjusting for PPP), twice the median non-AI salary of ₹125,000 (US$1,666).

The ‘AI wage premium’ persists, even after controlling for experience, education and other fixed effects. When including industry-region, industry-time and region-time fixed effects, we find that AI posts on average offer 30% higher wages than non-AI posts (see Model (1) of Figure 3.1). However, this may be driven by the highest-paying firms also disproportionately hiring AI roles. Therefore, we add firm fixed effects to control for differences between firms in Model (2). Even in this case, AI posts pay 19% more relative to the average non-AI post. Finally, posts that require AI skills may simply be different types of jobs. Models (3) and (4) therefore include fixed effects for the occupation and role, using respectively, the NCO 2004 classification codes and the more granular role label built into the online jobs site. A substantial AI premium of 13-17% remains.[14]

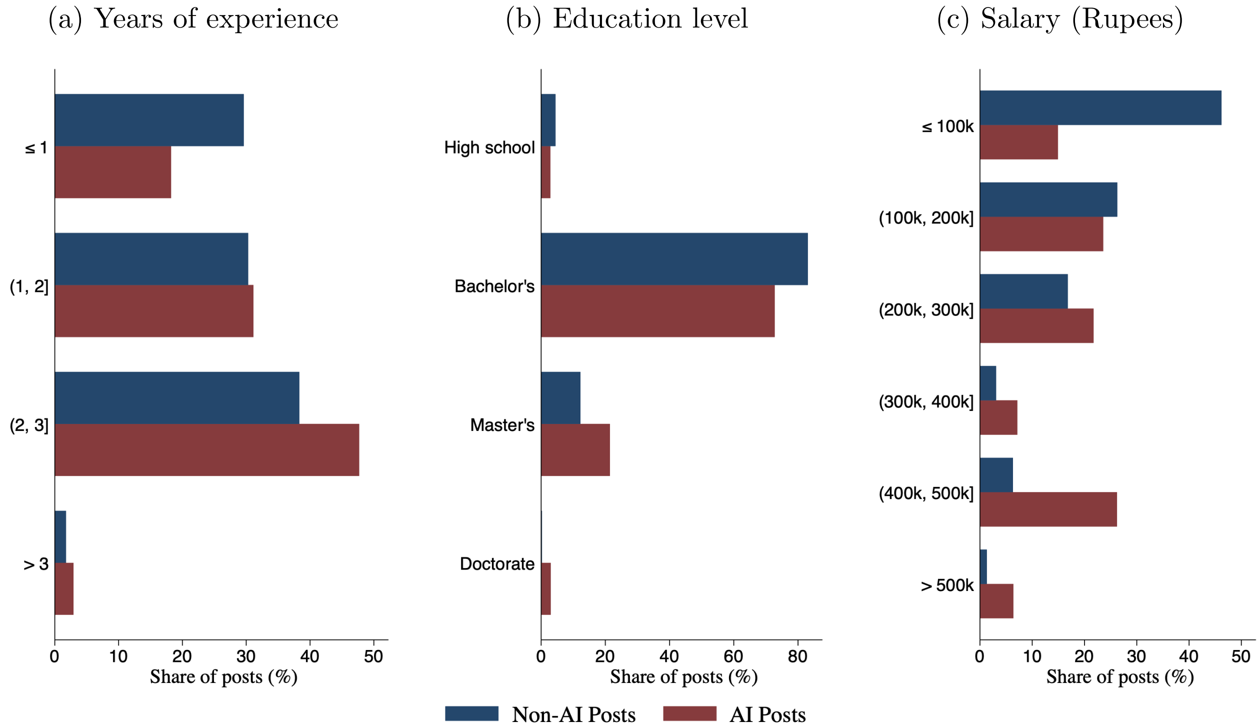
***3. AI roles are highly concentrated in a few key technology clusters, particularly Bangalore.***

AI demand is highly concentrated in large cities, particularly the major technology clusters around Bangalore, Mumbai, Hyderabad and Delhi (Figure 3.4). This reinforces the notion that AI adoption is occurring predominantly in the urban white-collar service sector. Bangalore alone has more than 30% of all AI vacancies across India. Panel (a) compares the shares of all posts across cities with their shares of AI posts, and shows that AI demand is even more spatially concentrated than hiring generally. The tail of all other cities are represented by the ‘Other’ category, which is significantly shorter for AI posts relative to the listed cities. Panel (b) shows the distribution of AI posts across districts. The vast majority of districts have few AI posts, since hiring is clustered in the largest cities. Shares of AI demand in cities have been remarkably constant over the last decade, as shown in Appendix Figure A.4, except for a prominent increase in AI activity in Mumbai as AI demand took off in the financial sector.

***4. AI roles are highly concentrated in the largest firms.***

[14] The interpretation of the control variables is as follows. An extra year of experience is associated with a more than 35% higher salary (at least within the predominantly early-career jobs posted on the site – see Figure 3.3), while having a Master’s degree is associated with up to 10% higher salary. In this sample, having only a high school education is associated with wage offers 3-6% below the baseline of having an undergraduate degree, though this figure is likely a dramatic underestimate of the effect, given the major under-representation of lower-skilled professions on the platform. The relationship between wage offers and having a doctoral degree is expressed predominantly through the firm- and role-effects: conditional on firm and occupation/role, there is no significant relationship to salary, but without such conditioning salaries are 7-13% higher. This is consistent with the wage offer premium for workers with doctorates being driven by taking higher-skilled jobs at more advanced firms.

Figure 3.3: Hiring profile of AI vs. non-AI vacancies




*Notes:* These graphs compare the distribution of posts, for AI and non-AI vacancies, across experience, education and salary. This information is reported directly in the online jobs platform. For experience and salary, the vacancy posts record a minimum and maximum value, so we take the midpoint of the specified range. AI posts are classified based on keywords, as described in Section 2.2.

Table 3.1: Wages in AI vs. non-AI roles

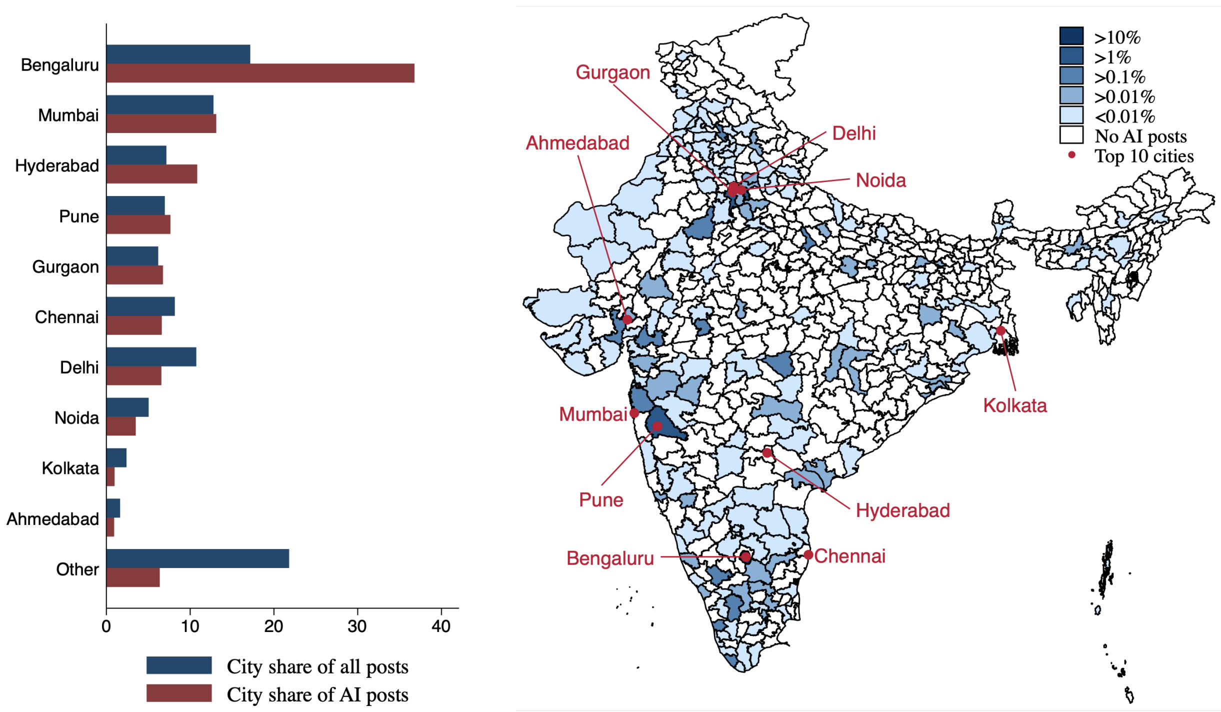
| | log Annual Salary | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| AI post | 0.308*** | 0.194*** | 0.131*** | 0.170*** |
| | (6.99) | (5.78) | (6.32) | (4.31) |
| Experience Required (Years) | 0.469*** | 0.411*** | 0.386*** | 0.351*** |
| | (68.46) | (52.10) | (48.95) | (43.49) |
| High School | 0.00519 | -0.0642*** | -0.0364** | -0.0396** |
| | (0.07) | (-3.37) | (-1.97) | (-2.31) |
| Master's | 0.103*** | 0.0772*** | 0.0414*** | 0.0199** |
| | (7.35) | (8.03) | (5.51) | (2.54) |
| Doctorate | 0.131** | 0.0719* | 0.00589 | 0.000334 |
| | (2.22) | (1.73) | (0.19) | (0.01) |
| *Fixed Effects:* | | | | |
| – Industry-Region | ✓ | ✓ | ✓ | ✓ |
| – Industry-Year | ✓ | ✓ | ✓ | ✓ |
| – Region-Year | ✓ | ✓ | ✓ | ✓ |
| – Firm | | ✓ | ✓ | ✓ |
| – Occupation Code | | | ✓ | |
| – Role Label | | | | ✓ |
| $R^2$ | 0.345 | 0.536 | 0.557 | 0.577 |
| Observations | 14080455 | 14044610 | 12933816 | 14044608 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. All regressions include industry-region, industry-time and region-time fixed effects, and models (2)-(4) also include firm fixed effects. *AI post* is a dummy such that the coefficient is the percentage increase in annual salary associated with posts requiring AI skills, after accounting for the control variables and fixed effects. Similarly, *Experience* is measured in years, so the coefficient reflects the percentage salary increase associated with an additional year of experience. The education variables are dummies, with the baseline category being a Bachelor's degree; for instance, *High School* reflects the percentage salary decrease associated with posts that only require a high school education. The *Occupation Code* fixed effect also accounts for variation across India's 4-digit National Classification of Occupations codes, while the more granular *Role Label* fixed effect accounts for variation across the self-selected role classifications built into the jobs portal.

Figure 3.4: Distribution of AI demand across cities

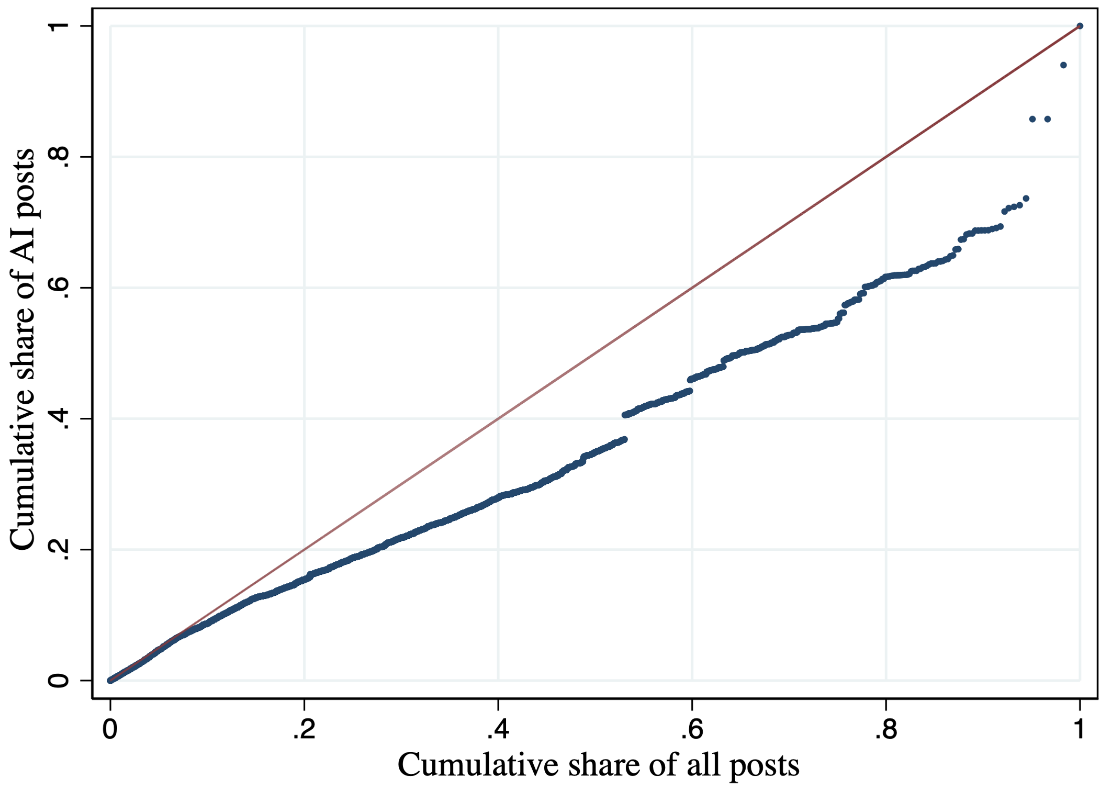
(a) Shares of posts across cities

(b) Share of all AI posts, by district, 2010-2019




*Notes:* The bar graph shows the shares of all posts and AI posts across cities, for the entire period 2010 to 2019. AI shares at the top end are larger than general post shares, indicating that online AI demand is more spatially concentrated than general hiring. The map shows the distribution of the share of all AI posts by particular districts. Labels are shown for the top ten cities with the most AI posts. The vast majority of districts have few AI posts, since hiring is clustered in the largest cities.

Figure 3.5: Distribution of AI posts across all firms, 2010-2019

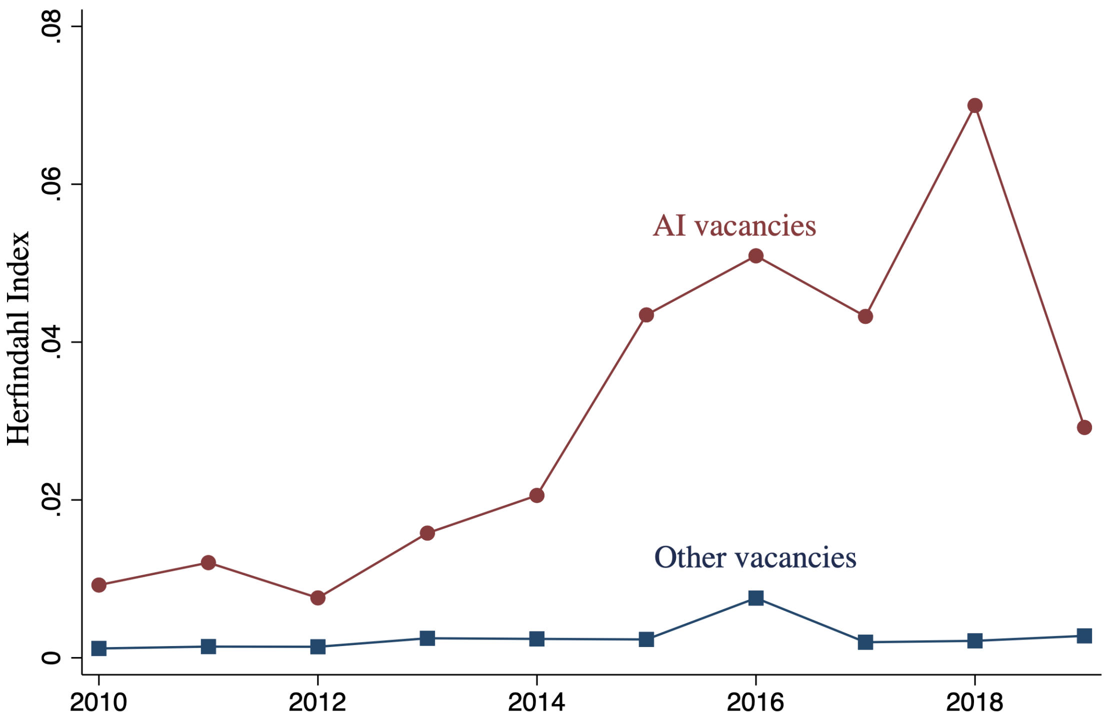



*Notes:* We plot the cumulative share of AI posts against the corresponding cumulative share of all posts. The red 45° line indicates a one-for-one increase in the share of AI posts relative to all posts. The deviation of our scatter plot from the 45° shows the extent to which AI vacancies are disproportionately posted by the largest firms.

Which firms hire AI skills? We proxy for firm size by the number of postings made on the platform. Figure 3.5 plots the cumulative share of AI posts against the corresponding cumulative share of all posts. This traces out a Lorenz-type curve, where the deviation from the 45° line shows the extent to which AI vacancies are disproportionately posted by the largest firms. Inspecting the top right corner reveals that the largest 14 firms are responsible for 10% of all vacancies, with each posting at least 50,000 vacancies, and these account for 31% of all AI posts. While there are some smaller firms that post a disproportionate number of AI posts, the largest AI-hiring firms are also the largest hirers in general.

To illustrate the increasing concentration of AI posts in the largest firms, Figure 3.6 plots the trend in the Herfindahl Index for AI and non-AI vacancies over time. In every year, AI posts are substantially more concentrated among the largest firms, with the gap widening rapidly from 2015. This implies that the take-off in AI demand in Figure 3.1 coincided with increased concentration in the hiring of AI skills, consistent with the notion that the larger firms are disproportionately more able to invest in the emergent deep learning technologies.

Figure 3.6: Firm concentration of AI posts, 2010-2019




*Notes:* We plot the trend in the Herfindahl Index for AI and non-AI vacancies over time. These are calculated for each year as the sum of squared firm market shares of all AI or non-AI posts, respectively.

***5. Having a first-mover AI adopter is associated with local AI diffusion, over and above industry and region trends, particularly in the IT sector.***

How did AI demand diffuse through the economy? The vast majority of industry-city pairs had zero postings seeking AI skills at the start of our period, so we can examine descriptively through an event study how other local firms started hiring AI skills after the first AI hire in a given city and industry. We first construct a time-varying dummy for each industry-city pair $ir$:

$$\text{FirstAI}_{irt} = \begin{cases} 1 \text{ in the first month where a firm } F \text{ in } ir \text{ posts an AI vacancy} \\ 0 \text{ otherwise} \end{cases}$$

We then construct the outcome AI $\text{share}_{irt,-F}$ by pooling all posts in the industry-city *except* those from the first adopter $F$. Combining these together, we run the event study specification:

$$\text{AI share}_{irt,-F} = \sum_{j=-2}^{2} \beta_{1j} \cdot \text{FirstAI}_{ir,t-j} + \alpha_{ir} + \alpha_{it} + \alpha_{rt} + \epsilon_{irt} \tag{3.1}$$

This gives the descriptive coefficients $\beta_{1j}$, which reflect the average percentage point increase in the AI share of vacancies posted in each year $j$ after the first adoption of AI in the city-industry pair. Crucially, this association is that which remains even after controlling for the broader industry- and city-level trends.

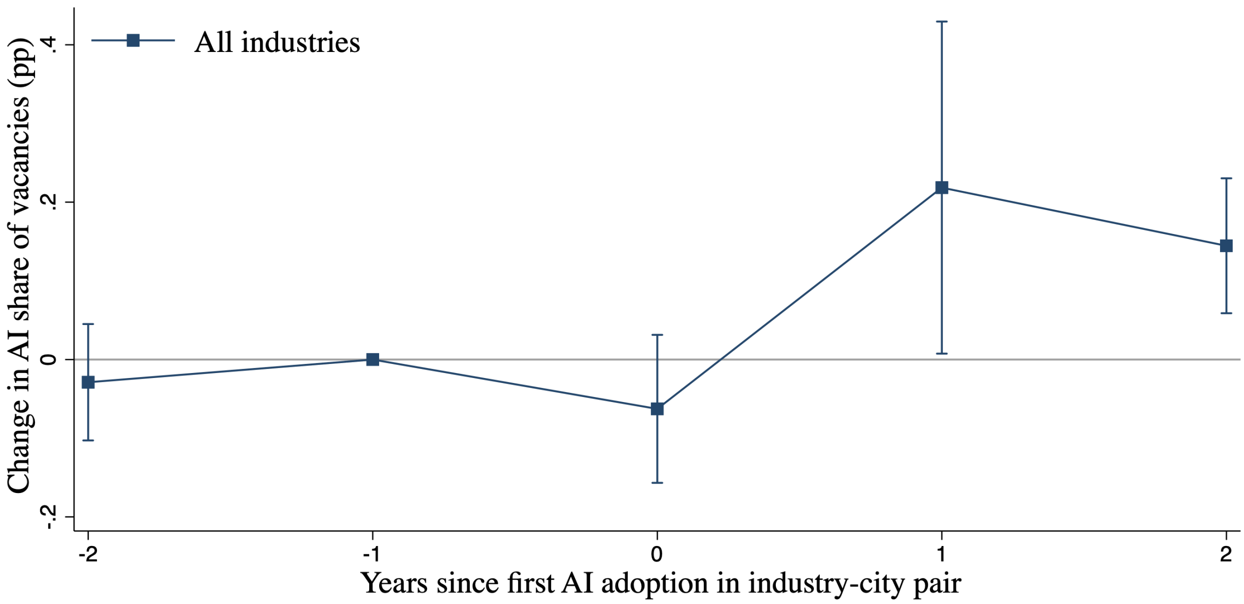
We find that there is a significant positive relationship between initial AI adoption and the share of AI postings by other local firms in subsequent years, as plotted by Figure 3.7 Panel (a). In the first year after the first AI post within an industry and city, the AI share is more than 0.2 percentage points higher ($p = 0.042$) than that in the absence of an AI adopter. This is a substantial difference considering that the average AI share of posts across all industries was only 1% by 2019 (see Figure 3.1). We also investigate heterogeneity in the diffusion of AI adoption across industries.[15] Results are shown in Figure 3.7 Panel (b) for the top five industries by AI share. There is substantial dispersion in the magnitude of the local effect, with by far the strongest relationship in IT & Software. We conclude that while local influence can be relevant for AI diffusion, its importance varies substantially across industries.

[15] Specifically, we estimate separate $\beta$s across industries $X_i$ by running the following specification:

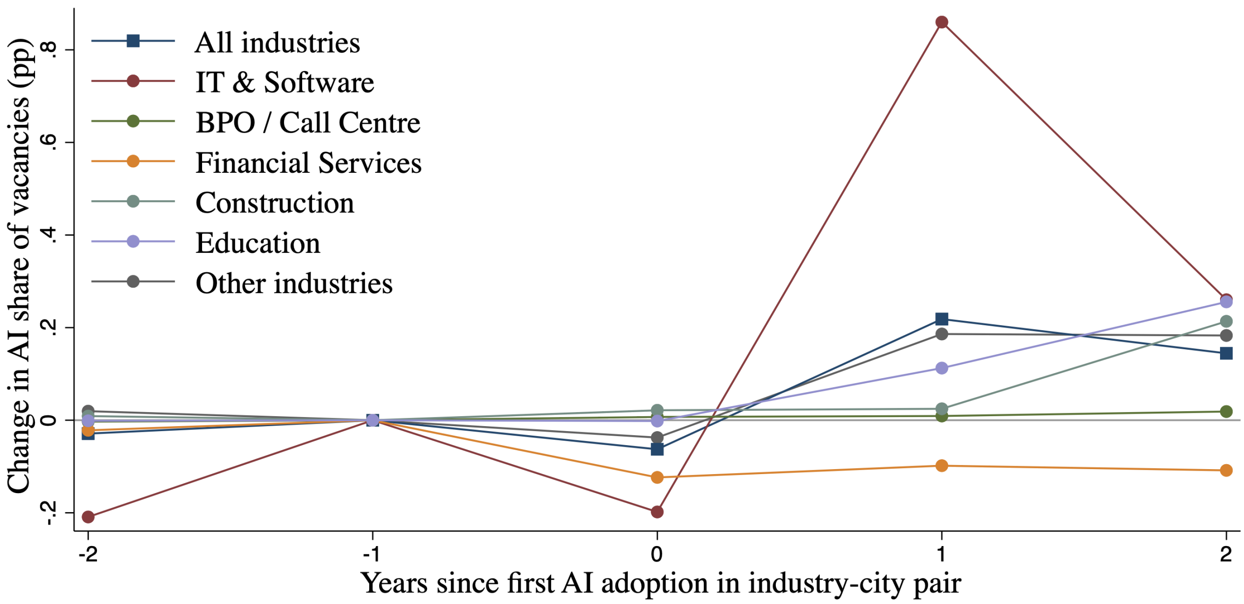
$$y_{irt,-F} = \sum_{j=-2}^{2} \beta_{2j} \cdot \text{FirstAI}_{ir,t-j} \cdot X_i + \alpha_{ir} + \alpha_{it} + \alpha_{rt} + \epsilon_{irt}$$

Figure 3.7: AI diffusion

(a) Change in AI share after first AI adoption in industry-city




(b) Heterogeneity across industries




*Notes:* Panel (a) shows the change in the share of vacancies that are AI vacancies, for years before and after the first adoption of AI in an industry-city pair. Panel (b) plots heterogeneity in the diffusion of AI adoption across industries. In all cases, AI posts by the initial AI adopter are excluded in order to focus on diffusion of AI posting to other firms in the industry-city pair.

# 4 Empirical Strategy

## 4.1 AI demand and establishment outcomes

The demand for AI skills ('AI demand') has been relatively widespread across occupations and industries and accounts for approximately 1% of all job postings in 2019, as discussed in Section 3. We now measure the effect of establishment-level AI demand on other hiring decisions.

Our three primary outcomes of interest are (1) the volume of AI posts, (2) the volume of non-AI posts, and (3) the wages offered by non-AI posts. We measure these outcomes as the long-difference growth between our baseline period, 2010-2012, and our endline period, 2017-2019, shortly after the take-off in AI demand in 2016.[16] Our primary unit of analysis is 'establishments', defined as firm-city pairs, which we use because many large firms have autonomous establishments in several different cities. Therefore, we estimate the effect of AI demand on the employment outcomes for a panel of almost 25,000 establishments together posting approximately two million vacancies on the platform within our baseline and endline periods.

We estimate the following empirical model for our main specifications in Section 5:

$$\Delta y_{fr,t-t_0} = \beta \cdot \Delta Adoption_{fr,t-t_0} + \alpha_r + \alpha_i + \alpha_{f10} + \epsilon_{fr,t-t_0}, \tag{4.1}$$

where $\Delta y_{fr,t-t_0}$ is the change in the inverse hyperbolic sine of outcome $Y_{fr}$, between 2010-2012 and 2017-19. $\Delta Adoption_{fr,t-t_0}$ is the change in the inverse hyperbolic sine of the number of AI posts by an establishment between 2010-12 and 2017-19. $\alpha_i$ and $\alpha_r$ are two-digit industry and city fixed effects. $\alpha_{f10}$ is a firm decile fixed effect, where firm deciles are calculated over the baseline period 2010-2012. We employ some or all of the fixed effects when reporting results in Section 5. Our preferred specification controls for region, industry and firm size fixed effects. Given that our unit of analysis are firm-city pairs, we cluster standard errors at the firm level. $\epsilon_{fr,t-t_0}$ is the mean zero error term.

The coefficient $\beta$ approximates the elasticity of an outcome with respect to AI demand: increasing the growth rate of AI demand by 1% between 2010-12 and 2017-19 causes a $\beta$ percentage point rise in the growth rate of the outcome variable across the same time period. Intuitively, the key variables $\Delta y_{fr,t-t_0}$ and $\Delta Adoption_{fr,t-t_0}$ are (approximately) the growth in establishment outcomes and AI demand between 2010-12 and 2017-19.[17]

[16] We pool within these periods in order to improve precision and maximise the probability that a firm advertises on the job postings platform during both time periods.

[17] Mathematically, for growth rate $g$ defined by $Y_t = (1+g)Y_{t_0}$, and using the approximation that $\ln(1+g) \approx g$

## 4.2 Instrumenting AI demand by AI exposure

When considering the relationship between AI demand and establishment-level outcomes, we note that AI demand is likely to be endogenous, reflecting unobserved differences in establishment productivity.[18] Therefore, we instrument AI demand by 'AI exposure' to isolate supply-side technical advances in the capabilities of AI and employ a two-stage least squares model. Specifically, we use an AI exposure measure developed by Webb (2020) that captures the degree of overlap between workers' tasks and tasks that can be performed by patented AI technologies. Specifically, Webb (2020) measures the overlap between the text of AI patents and the text of O*NET job task descriptions.[19] Occupations with a greater overlap in tasks that are capable to be automated by AI are assigned a higher exposure measure.

We use publicly-available crosswalks to map the Webb (2020) exposure measure to the Indian National Classification of Occupations (NCO) 2004 at the four-digit level (see Appendix B for more detail). By aggregating this measure up to the establishment level using baseline establishment occupation shares, we capture plausibly exogenous variation in the innate technological feasibility of using AI in business activities. We can also aggregate this exposure score to an industry or district level (see Figure A.5). In Figure 4.1, we map the exposure score across the occupational wage offer distribution and find that AI exposure rises with wage offers up to a peak around the 80th percentile, before falling thereafter.[20]

To speak to the validity of this instrument, we note that India is not a significant producer of new AI research and lags far behind other major AI research hubs on per capita terms, predominantly the USA, China and Singapore.[21] In our analysis, we drop vacancy posts from AI-producing sectors in order to focus on AI-using sectors.[22]

for small $g$, we have $g = \ln Y_t - \ln Y_{t_0} = \Delta y_{t-t_0}$.

[18] For example, more innovative managers are more likely to hire more AI workers, but they are also more productive and grow the business more quickly, increasing labour demand in general.

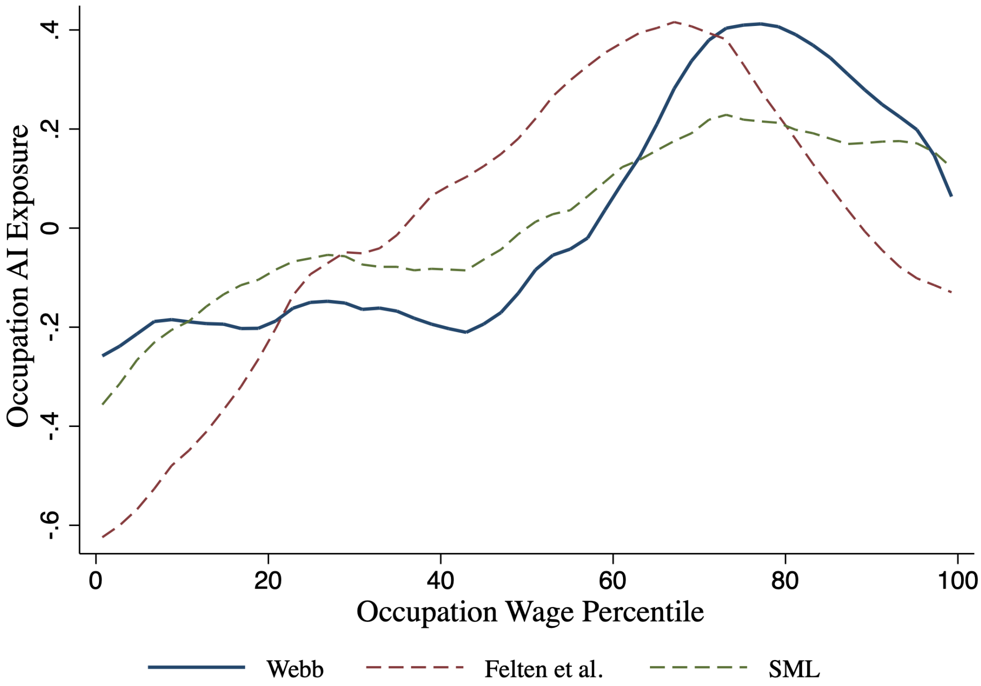
[19] These task descriptions are based on US occupations. Whilst Indian occupations in general may have very different task compositions, the white-collar service sector is relatively similar. To the extent that this is not the case, it would just count against the strength of our first stage – a further advantage of our two-stage least squares approach over simply correlating exposure with outcomes.

[20] Intuitively, many low-paid services jobs require elements of manual dexterity or social interaction which are hard to automate (e.g. waiters or school-teachers), then there is a middle range (e.g. insurance brokers, legal associates) which are highly exposed, while the top end (e.g. CEOs) are again hard to automate.

[21] Despite strengths in applied computer science and engineering founded on the Indian Institutes of Technology, India is not a significant producer of new AI patents (Perrault et al. 2019). Figure A.2 ranks each country on a wide range of AI progress metrics. In terms of the number of AI patents, the USA is dominant. Similarly, the USA, China and Singapore are also significant producers of AI conference papers and journal articles.

[22] These sectors include education, IT, internet and e-commerce, telecom and internet service providers, which make up 34.8% of our sample.

Figure 4.1: AI exposure by occupation wage offers




*Notes:* This graph shows a smoothed local polynomial regression of the Webb AI exposure measure on occupational wage offers. We first rank occupations by their average salary across all vacancy posts 2010-2019. We then plot the AI exposure associated with each, smoothing across a bandwidth 10 of percentage points. In addition to our main measure, from Webb (2020), we also show analogous results for the alternative measures (Felten et al. 2018, Mani et al. 2020) which we use in robustness checks in Section 6.

We estimate the following first stage in our main specifications:

$$\Delta Adoption_{fr,t-t_0} = \gamma \cdot Exposure_{fr,t_0} + \alpha_r + \alpha_i + \alpha_{f10} + \epsilon_{fr,t-t_0}, \tag{4.2}$$

where we instrument $Adoption_{fr,t-t_0}$ with baseline AI exposure $Exposure_{fr,t_0}$, constructed as a weighted average of the AI exposure measure of Webb (2020). Specifically, we construct occupation-level exposure as the interaction between the share of posts and exposure to AI adoption based on US patents in 2010-2012, as follows:

$$Exposure_{fr,t_0} = \sum_o PostShare_{fro}^{t_0} \cdot ExposureMeasure_o \tag{4.3}$$

$Exposure_{fr,t_0}$ is standardized to have a mean of zero and a standard-deviation of one. Therefore, the first stage coefficient $\gamma$ in Equation 4.2 approximates the proportional change in AI posts between 2010-12 and 2017-19 that is associated with a one standard deviation rise in AI exposure.

# 5 Results

We run a two-stage least squares model to estimate the effect of AI vacancy growth on the growth of establishment-level hiring outcomes for our panel of vacancy-posting firms. We first document the first stage, and show that firms more exposed to AI significantly increase their growth in AI demand over time. We then turn to the second stage to examine the impact of the exogenous growth in AI demand on the change in establishment-level non-AI vacancies and wage offers over time.

## 5.1 AI exposure predicts AI demand

Firms with higher AI exposure scores are differentially increasing demand for AI skills, using the Webb (2020) exposure measure. A one standard deviation rise in the baseline AI exposure score at the establishment level is associated with a 1.93% statistically significant increase ($p<0.01$) in the number of AI vacancies between 2010-12 and 2017-19. Table 5.1 summarises the first-stage results and shows that the relationship holds after controlling for region, firm size and industry fixed effects (Column 2). The result is also robust to controlling for region and firm size (Column 1) and region and firm fixed effects (Column 3), albeit with a weaker relationship (0.06%, $p<0.05$), which is unsurprising given that the variation is derived solely across establishments within firms. We conclude that AI exposure is a relevant instrument for the growth in demand for AI skills.

Table 5.1: First stage: Impact of AI exposure on establishment AI demand

| | Growth in AI Vacancies | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Establishment AI Exposure | 0.0170*** | 0.0193*** | 0.00607** |
| | (5.13) | (5.21) | (2.05) |
| *Fixed Effects:* | | | |
| – Region | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | |
| – Industry | | ✓ | |
| – Firm | | | ✓ |
| $R^2$ | .0341 | .049 | .3774 |
| Observations | 22,251 | 22,251 | 19,383 |

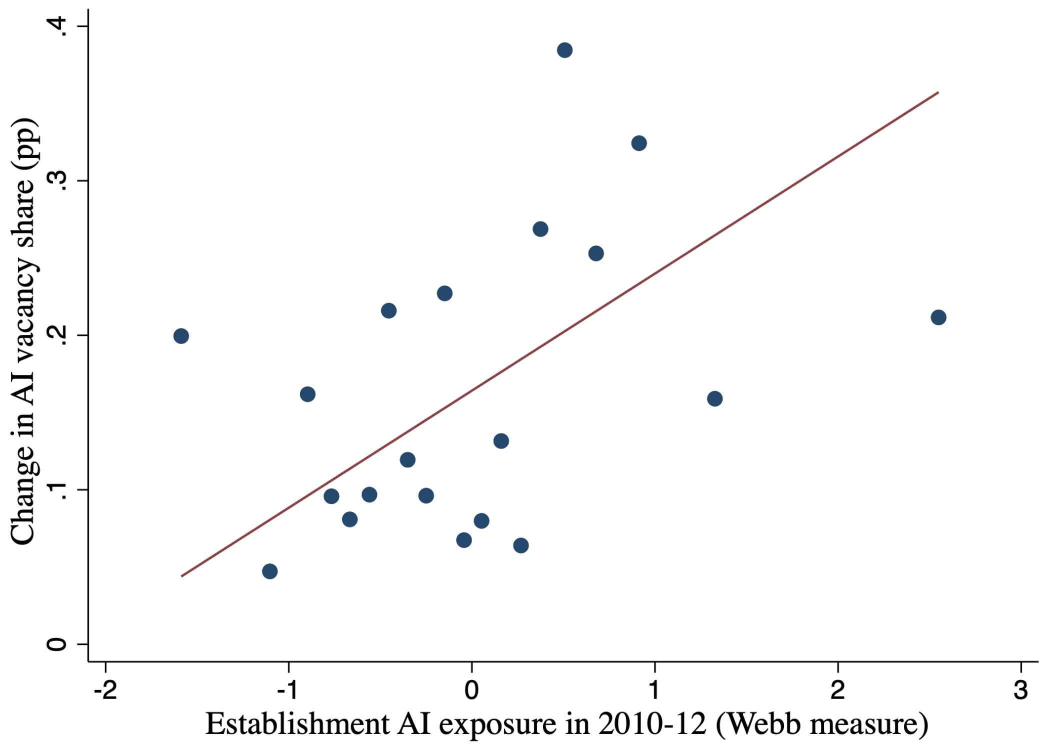
*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. The dependent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. The independent variable is establishment AI exposure, calculated as the standardized average of occupation AI exposure (from Webb 2020), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020). Each coefficient therefore represents the proportional impact on AI hiring of a one-standard deviation rise in AI exposure.

The relationship also holds when examining the effect of AI exposure on the AI share of vacancies at establishment level, both in long differences (Figure 5.1 Panel (a)) and by quantile of AI exposure and year (Figure 5.1 Panel (b)). The most exposed establishments have the highest AI share of vacancies (almost 8% in 2019), driving most of the take-off in AI demand after 2016. However, we focus on the growth in the number of AI vacancies to avoid spurious correlations.[23]

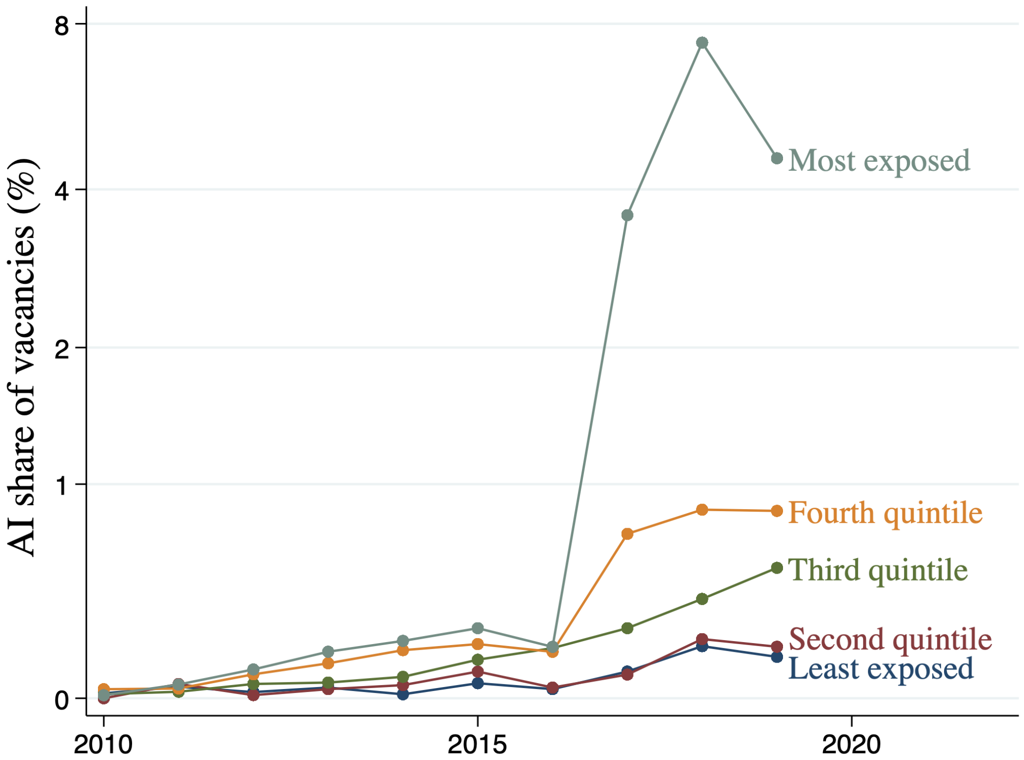
[23]Regressing total posts on the AI share (AI posts over total posts) would likely have a mechanical negative relationship, as demand shocks for non-AI workers would affect the denominator of AI share and the outcome variable. By using the growth in the number of AI posts as the independent variable, such demand shocks instead count against our story: a positive demand shock for AI workers raises the number of AI posts, and the number of total posts. Thus, our findings of a negative impact of AI demand growth on the change of total establishment employment over time are if anything an under-estimate.

Figure 5.1: Impact of AI exposure on establishments' AI share of posts

(a) Long differenced AI share vs. exposure




(b) Annual AI share by exposure quintile




*Notes:* These graphs show the relationship between AI exposure and establishments' shares of AI posts. The binned scatter plot in (a) summarizes the relationship between baseline AI exposure and establishments' change in AI vacancy share between 2010-12 and 2017-19. The covariates from column (2) of Table 5.1 are partialled out. Panel (b) plots the time variation in this relationship, using an inverse hyperbolic sine scale for the y-axis.

## 5.2 AI demand lowers the growth in non-AI demand

We now turn to the second stage to examine the effect of AI demand on non-AI vacancies and wage offers. We note that these outcomes pertain to new hires that firms intend to make, rather than those of existing workers. Table 5.2 shows the effect of the growth of AI vacancies on the growth of non-AI vacancies and total vacancies, instrumenting on AI exposure using the Webb (2020) exposure measure. The growth in AI demand reduces the growth in non-AI hiring intent: taking column (2) as our baseline result, 1% increase in the growth rate of AI vacancies results in a 3.6 percentage point decrease ($p$<0.01) in the change in non-AI vacancies at establishment level between 2010-12 and 2017-19, controlling for region, firm size and industry fixed effects. There is a similarly sized decrease of 3.57 percentage points in the growth rate of total vacancies, highlighting that the growth in AI vacancies disproportionately crowds out growth in other vacancies in the white-collar service sector. Importantly, we examine this relationship in the years directly following the sharp take-off in AI demand. This negative relationship is consistent with the findings of Acemoglu et al. (2020), who show that AI exposure (different to our result of AI demand itself) results in lower non-AI hiring in the USA. Both results are robust to controlling for region and firm size fixed effects separately (Column 1), and to controlling for region and firm fixed effects (Column 3). However, the latter result is only marginally statistically significant at the 10% level and the instrument is weak, given the limited variation across establishments within firms.

Table 5.2: Second stage: Impact of AI demand on the growth of non-AI vacancies

| | Growth in Non-AI Vacancies | | | Growth in Total Vacancies | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -5.942*** | -3.605*** | -9.944* | -5.909*** | -3.566*** | -9.923* |
| | (-3.66) | (-3.16) | (-1.84) | (-3.64) | (-3.14) | (-1.84) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 26.31 | 27.17 | 4.185 | 26.31 | 27.17 | 4.185 |
| Observations | 22,251 | 22,251 | 19,383 | 22,251 | 22,251 | 19,383 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Webb 2020), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

## 5.3 AI demand lowers the growth in non-AI wage offers

Does increased AI demand reduce the growth in wage offers for non-AI vacancies? We exploit the fact that firms use a standard template to advertise vacancies on our job postings platform, providing us with wage offer data for all job adverts. Table 5.3 documents the effect of the growth in AI vacancies on the growth of median wages for non-AI postings and all job postings.

We find that a 1% higher growth rate in AI vacancies between 2010-12 and 2017-19 reduces the growth rate of non-AI wage offers by 2.6 percentage points ($p$<0.01) across the same time period, instrumenting on AI exposure and controlling for region, firm size and industry (Column 2). As with vacancy growth, the negative effects of AI demand are hardly changed when considering all posts, inclusive of AI postings. We observe a similar decrease of 2.5 percentage points ($p$<0.01) in the growth rate for the median wage offer of all job postings (Column 5). Again, our results are robust to alternative use of fixed effects, with the same caveats for firm fixed effects.

The reduction in the growth rate of non-AI wage offers in response to increased AI demand occurs across the entire wage offer distribution, with statistically strong effects after the 20th percentile and imprecisely estimated effects at the lowest decile. Figure 5.2 illustrates the percentage point impact of a 1% higher growth rate in AI demand on the growth rate of a given percentile of the wage offer distribution at establishment level between 2010-12 and 2017-19,

Table 5.3: Second stage: Impact of AI demand on the growth of non-AI wages

| | Growth in Non-AI Median Wage | | | Growth in Overall Median Wage | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -3.101*** | -2.599*** | -5.973* | -3.017*** | -2.527*** | -5.696* |
| | (-3.47) | (-3.43) | (-1.83) | (-3.50) | (-3.46) | (-1.87) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 25.64 | 26.39 | 4.294 | 26.84 | 27.71 | 4.602 |
| Observations | 22,064 | 22,064 | 19,217 | 22,071 | 22,071 | 19,223 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure, over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

instrumented with AI exposure and controlling for region, firm size and industry fixed effects. Across the wage offer distribution from the 20th percentile onwards, we observe a statistically significant reduction in wage offers for non-AI jobs over time, ranging from -1.3 to 2 percentage points ($p$<0.05). Even the reduction in the growth of wage offers at the 10th percentile is weakly statistically significant, at the 10% level, with a $t$-statistic of -1.92.[24][25] Mid-wage offers are most affected by the changes in wage growth over time, although these results are not statistically significantly different from the changes to low and high-wage offers.[26]

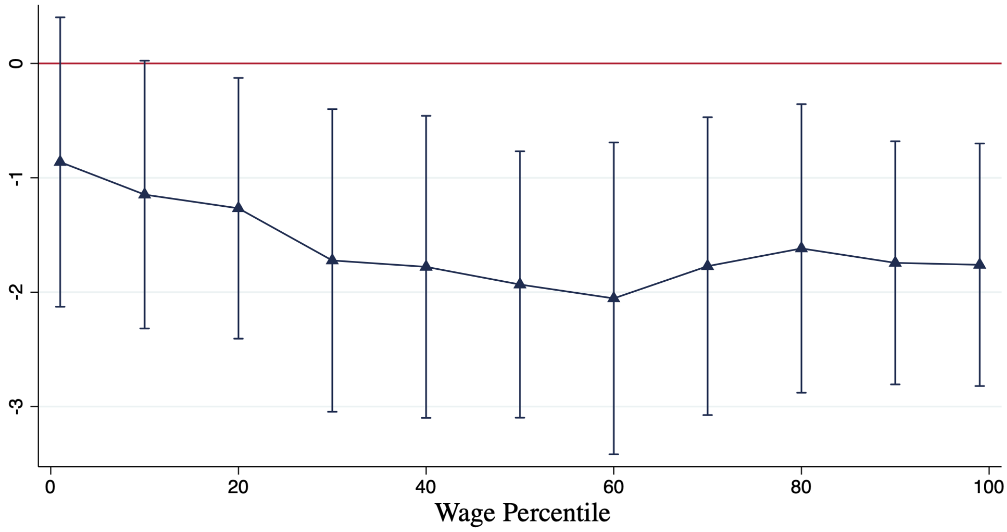
[24] The negative effect on the wage growth at the 1st percentile of the wage offer distribution is the only non-statistically significant result at conventional levels of significance.

[25] While we observe a fall in wage offers for each decile over time, we remain agnostic about the change in the composition of jobs at each decile.

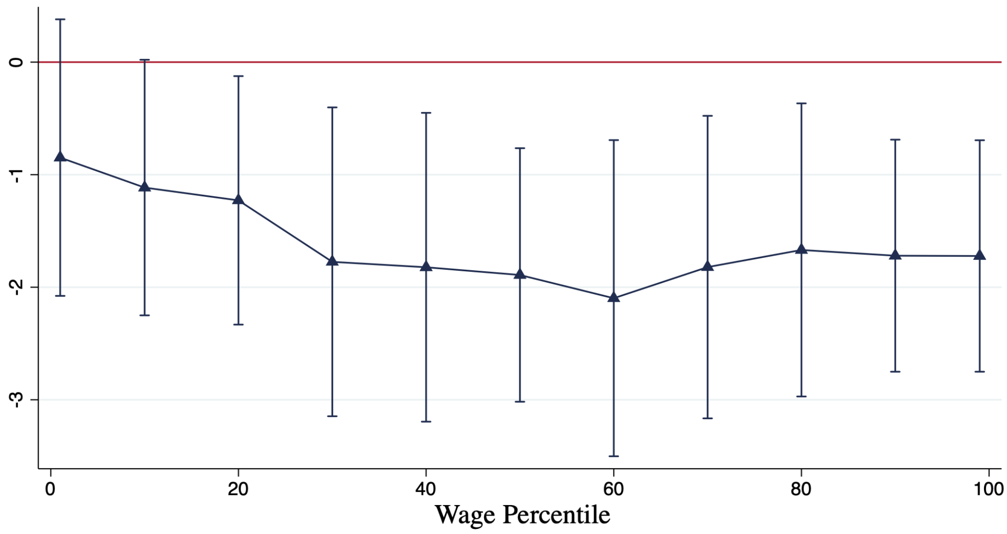
[26] This pattern parallels a long literature on ‘hollowing out’ (Goos & Manning 2007), as well as more recent findings focused on the impact of industrial robotics (e.g. Dixon et al. 2019).

Figure 5.2: Impact of AI demand on the wage offer distribution

(a) Wage growth in Non-AI posts only




(b) Wage growth in all posts




*Notes:* These coefficient plots show the impact of increased establishment AI demand on wage growth over time across the distribution of establishment wage offers. Each coefficient in Panel (a) is from a regression of type (2) in Table 5.4, and likewise each coefficient in Panel (b) is from a regression of type (5). In other words, each coefficient represents the percentage point impact of a 1% higher growth in establishment AI demand on wage growth over time for a given percentile of the wage offer distribution. We report the 1st and 99th percentile of the wage offer distribution and deciles in between the two extremes, alongside 95% confidence intervals. As in Table 5.4, AI demand is instrumented with AI exposure, standard errors are clustered at the firm level, and we include region, firm decile and industry fixed effects. Since AI posts make up only a small share of all roles in most establishments, the pattern is very similar across the two distributions.

## 5.4 Mechanisms

These results for wage offers could reflect one of two possibilities. The first is that firms are putting out lower wage offers for the same roles. The second is that firms are simply hiring different types of workers. We therefore also explore whether these wage offer results hold when additionally controlling for changes in the job requirements specified in the vacancy posts.

We find that the reduction in the growth of non-AI wage offers persists, even when controlling for changes in education and experience requirements over time. Table 5.4 shows the residual effect of the growth of AI vacancies on the growth of the median wage offers of non-AI posts and all posts, instrumenting on AI exposure and after controlling for the growth in the average experience and education levels. Even when controlling for changes in job profiles over time, a 1% higher growth rate in AI vacancies reduces the growth rate in the median wage offers of non-AI posts by 1.93 percentage points (Column 2) and of all posts by 1.89 percentage points (Column 5) between 2010-12 and 2017-19, both precisely estimated at the 1% level of significance. These coefficients are, however, slightly lower than those in Table 5.3, suggesting that both potential explanations play a role. Greater AI demand both changes the type of workers hired, and reduces the wage offer conditional upon worker profiles.

We can also see the impact on the type of workers hired directly, by examining how AI demand affects average education or experience requirements. Table 5.5 shows the results: higher growth in AI demand reduces the growth rate in the years of experience demanded by establishments in their non-AI postings, but has no statistically significant effect on the share of vacancies requiring a postgraduate degree.

We also dis-aggregate the effects across different types of worker. Figure A.6 in Appendix A plots the impact of increased establishment AI demand on the growth in median wage offers, within broad occupational groups. Specifically, we repeat models (2) and (5) from Table 5.3, except that in each case we construct the outcome variable from establishment wage offers within just one occupational category. While the results are not statistically significant due to the smaller sizes of the sub-samples, they suggest that it is high-skilled occupations that most lose out from establishment AI adoption.[27] This aligns with the findings of Webb (2020), who notes that the occupations most exposed to AI – i.e. those involving pattern-detection, judgement and optimization, such as clinical laboratory technicians, chemical engineers, optometrists, and power plant operators – are disproportionately high-skilled jobs.

[27]This qualitative finding is also robust to examining mean wages, as in Appendix Figure A.7.

Table 5.4: Second stage: Impact of AI demand on the growth of non-AI wages, controlling for changes in job profiles

| | Growth in Non-AI Median Wage | | | Growth in Overall Median Wage | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -2.132*** | -1.933*** | -5.340* | -2.088*** | -1.891*** | -5.103* |
| | (-3.16) | (-3.25) | (-1.84) | (-3.20) | (-3.29) | (-1.89) |
| Growth in Experience (Years) | 0.836*** | 0.824*** | 0.708*** | 0.836*** | 0.823*** | 0.711*** |
| | (27.95) | (29.03) | (15.46) | (28.10) | (29.15) | (16.21) |
| Growth in High School share | -0.0662 | -0.0830 | -0.177** | -0.0692 | -0.0860 | -0.179** |
| | (-0.73) | (-0.98) | (-2.03) | (-0.78) | (-1.04) | (-2.10) |
| Growth in Master's share | 0.254*** | 0.257*** | 0.282*** | 0.252*** | 0.255*** | 0.280*** |
| | (7.15) | (7.21) | (3.85) | (7.11) | (7.18) | (3.94) |
| Growth in Doctorate share | 2.669** | 2.385** | 3.549 | 2.624** | 2.345** | 3.384 |
| | (2.08) | (2.14) | (1.26) | (2.09) | (2.15) | (1.27) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 26.1 | 26.84 | 4.21 | 27.31 | 28.16 | 4.522 |
| Observations | 22,064 | 22,064 | 19,217 | 22,071 | 22,071 | 19,223 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure, over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

Table 5.5: Impact of AI demand on establishment non-AI education and experience

| | Growth in Non-AI Postgraduate Vacancy Share | | | Growth in Non-AI Years of Experience | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -0.225 | -0.319 | -0.573 | -1.065*** | -0.691** | -0.705 |
| | (-1.00) | (-1.55) | (-0.76) | (-2.81) | (-2.29) | (-0.72) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 25.12 | 25.87 | 3.894 | 25.12 | 25.87 | 3.894 |
| Observations | 22,244 | 22,244 | 19,377 | 22,244 | 22,244 | 19,377 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Webb 2020), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020). *Non-AI Postgraduate Vacancy Share* is defined as the establishment-level share of non-AI posts requiring either a Master's or a Doctorate.

### 5.5 Wider effects

Thus far, we have examined the effect of AI demand on establishment-level job posting outcomes. How does AI demand affect broader employment outcomes within industries and districts? We aggregate our vacancy data to industry-district level and find that AI exposure does predict AI demand (Table A.12 columns (1) and (2)). A one standard deviation increase in AI exposure is associated with a 1.1% increase in AI vacancies, controlling for industry and district fixed effects. However, the first stage is not sufficiently strong to detect significant effects on the growth of non-AI vacancies or wage offers (columns (3) to (6)). One interpretation of these results could be that while improvements in AI technologies have indeed spurred AI demand within some services industries, the demand is not yet sufficiently widespread for general effects to appear.

## 6 Robustness

In this section, we demonstrate that our establishment-level results are robust to alternative AI exposure measures and model specifications. We also find support for our results in other administrative datasets.

## 6.1 Alternative exposure measures

For our main specifications, we use the AI exposure measure proposed by Webb (2020), as it measures which tasks overlap with capabilities outlined in AI patents, thereby objectively capturing supply-side advancements in the AI technological frontier. Webb (2020) also validates the measure against previous IT and robotic trends. However, alternative AI exposure measures have also been proposed in the literature to date. Therefore, we examine whether our results remain robust to alternative definitions of our instrument.

We first consider the AI exposure measure proposed by Felten et al. (2018). Their AI Occupational Impact measure draws on data from the AI Progress Measurement project from the Electronic Frontier Foundation. The data identify nine application areas in which AI has made progress since 2010. Felten et al. (2018) crowdsource assessments on the applicability of these applications to 52 O*NET ability scales using Amazon MTurk. The AI Occupational Impact assigns an AI exposure score to each O*NET occupation as the weighed sum of the 52 O*NET ability assessments, where the weights are equal to the O*NET-reported prevalence and importance of each ability within each occupation. We map the Felten et al. (2018) measure to Indian NCO using a publicly available crosswalk (see Appendix B).

Our results remain robust to the use of the Felten et al. (2018) AI exposure instrument. We first observe that the AI exposure predicts AI demand in the first stage (Table A.1). Turning to the second stage, we observe that the negative effects on the growth of wage offers in response to increased AI demand remains robust to the use of Felten et al. (2018) as an instrument. A 1% higher growth in AI demand results in a 1.51% decrease ($p$<0.05) in the growth rate in wage offers between 2010-12 and 2017-19 (Table A.3). This result strengthens to 1.95% reduction ($p$<0.01) in the growth rate for the non-AI median wage offer, after controlling for changes in the education and experience requirements over time (Table A.4). Moreover, we similarly observe a negative effect on the growth rate across the entire wage offer distribution, except for the very lowest percentiles (Figure A.8). However, we do not observe any significant effects on the growth of non-AI vacancies (Table A.2).

We also consider the Suitability for Machine Learning (SML) methodology from Brynjolfsson et al. (2018), which uses surveys to score O*NET direct work activities against a rubric of suitability for machine learning (e.g. inputs and outputs are machine-readable, feedback is immediate, task is principally concerned with matching or prediction, etc.). We use an India-specific version of the SML index created by Mani et al. (2020), who interviewed more than 3000 Indian employees using the SML rubric and mapped a SML score onto every occupation in the 2004 NCO at the four-digit level. However, the SML exposure measure fails to predict firm demand for AI skills using our job vacancy data. In fact, we find that establishments scoring high on the SML measure are actually *less* likely to adopt AI over the period, in contrast to the other two measures (Tables A.1 and 5.1). One explanation for these differences could be that the Webb (2020) measure is based on current patented technological capabilities, whereas the SML measure is more forward-looking in its predictions.

## 6.2 Alternative specifications

The wage results are robust to using mean rather than median wage offers (Table A.5). Our results are also robust to weighting by baseline establishment size, proxied by number of job postings advertised between 2010 and 2012, with the top 5% winsorised (Tables A.6, A.7 and A.8). The results are also qualitatively robust to weighting by establishment size even when winsorising only the top 0.5% (approximately 100 firms) (Tables A.9, A.10 and A.11). The median size of these 100 largest firms is 200 times that of the median firm overall, with the very largest firm almost 1500 times the size of the overall median firm. Thus, these very largest firms dominate the distribution when the full set of weights are used, and the first stage becomes very weak.

## 6.3 Alternative data sources

As an additional robustness check, we investigate potential wage effects of AI exposure in the 2011-12 National Sample Survey (NSS) and 2017–18 Periodic Labour Force Survey (PLFS) administrative datasets, which are nationally representative labour surveys. As noted in Section B.2, we cannot observe AI adoption in these datasets. However, we can use the baseline industry-region occupation distribution to assess AI exposure, and hence examine reduced form relationships. Table A.13 thus presents results on the relationship between AI exposure and wage growth in NSS/PLFS industry-districts.

Columns 3-6 show the relationship between AI exposure and the growth between 2011-12 and 2017-18 in industry-district-level average wages, using these surveys. In this case, we can measure exposure at the industry-district level, and find a strong negative relationship between AI exposure and wage growth. Specifically, having one standard deviation higher AI exposure is associated with 7.7% lower growth in median wages over the period. While this evidence lacks the strength of our IV approach, which leverages observed AI adoption from the vacancy posts, it is consistent with a negative impact of AI adoption on the wages of workers in services firms.[28]

# 7 Conclusion

In this paper, we use a novel dataset of online vacancy posts from India's largest jobs website to shed light on the demand for AI skills in the service sector and the subsequent impacts on establishment-level labour demand. We firstly show that there was a rapid take-off in AI demand after 2016, particularly in the IT, finance and professional services industries – aligning with trends found in the USA and UK (Acemoglu et al. 2020, Stapleton & O'Kane 2020). AI roles attract a substantial wage premium of 13 to 17%, strikingly similar to that documented in the USA by Alekseeva et al. (2019). AI vacancies are also highly concentrated in the largest firms and a few key technology clusters. These descriptive results are consistent with rapid global diffusion of AI capabilities, at least to those large and high-performing firms close to the technological frontier.

We next investigate the effects of growth in establishment AI demand between 2010 and 2019 on other establishment-level labour outcomes. To isolate causation, we exploit establishment-level variation in exposure to supply-side advances in AI capabilities, aggregated up from the occupational exposure measure of Webb (2020), as an instrument for AI demand. In the first stage, establishments that were *ex ante* more exposed to advances in AI significantly increase demand for AI skills in their online vacancy posts. In the second stage, we find that the growth in AI demand has a significant negative effect on the growth in other non-AI postings by establishments. Specifically, a 1% increase in the AI vacancy growth rate results in a 3.6 percentage point decrease in the growth of non-AI vacancies within establishments between 2010-12 and 2017-19, controlling for region, firm size and industry fixed effects. These results suggests that, at the establishment level, the displacement effects of AI appear to outweigh any productivity or labour reinstating effects, consistent with findings in the USA by Acemoglu et al. (2019).

Finally, we show that growth in AI demand also reduces growth in the average wage offers of

[28] Columns 1-2 show the relationship between baseline AI exposure and the growth between 2011-12 and 2017-18 in large services firms' wage bills, using Prowess. Given that there is no statutory obligation to report the number of employees, and fewer than 10% do so in practice, we focus on impacts on the total wage bill; we see this measure as a composite of our hiring and wage margins. As noted in Section B.2, Prowess only contains large and predominantly publicly-listed firms; hence, the data contains a comparatively small number of observations and we can aggregate only to the industry-state level. We estimate baseline exposure by combining the Webb (2020) measure with the industry-state distribution of white-collar services occupations in the 2011-12 NSS. With these data limitations in mind, we find a negative relationship that is consistent with our other findings, but we cannot reject that this effect is statistically significant different from zero.

the vacancies posted, both for non-AI roles only and overall. We find that a 1% higher growth rate in AI vacancies between 2010-12 and 2017-19 reduces the growth rate of non-AI wage offers by 2.6 percentage points, instrumenting on AI exposure and controlling for region, firm size and industry. This negative effect appears across the wage offer distribution after the 20th percentile. We further find that these wage offer results hold for the mean as well as median wage offer, and after additionally controlling for changes in education and experience requirements of postings. These results suggest that the drop in the wage offer growth rate cannot be attributed solely to firms hiring lower-skilled workers in response to AI adoption. Furthermore, our results are robust to using alternative AI exposure measures and labour force survey data at the regional level.

Taken together, our findings highlight the 'double-edged' impacts of AI, aligning with existing literature from developed countries. AI jobs pay a substantial wage premium, but these opportunities are highly concentrated in certain industries, cities and large firms. AI adoption within an establishment reduces both the number of other job vacancies posted on the platform and the corresponding wage offers. Such net displacement effects within the firm could have important negative consequences if they are not balanced out by positive effects elsewhere in the economy.

However, in this paper we have only explored the net direct effect of adopting AI on pre-existing establishments. We do not take a stand on whether this AI adoption has had positive effects on other firms, or has created new jobs elsewhere in the economy. Tracing these potential 'creative' effects, and evaluating whether they offset our 'destructive' effects, are important tasks for future research.

# References


Acemoglu, D., Autor, D., Hazell, J. & Restrepo, P. (2019), 'AI and Jobs: Evidence from Online Vacancies', p. 59.

Acemoglu, D., Autor, D., Hazell, J. & Restrepo, P. (2020), 'AI and Jobs: Evidence from Online Vacancies', p. 51.

Acemoglu, D. & Restrepo, P. (2018), 'The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment', *American Economic Review* **108**(6), 1488–1542.

Acemoglu, D. & Restrepo, P. (2019), The Wrong Kind of AI? Artificial Intelligence and the Future of Labor Demand, Technical Report w25682, National Bureau of Economic Research.

Adams, A., Balgova, M. & Qian, M. (2020), 'Flexible Work Arrangements in Low Wage Jobs: Evidence from Job Vacancy Data', p. 46.

Agrawal, A., Gans, J. & Goldfarb, A. (2018), *Prediction Machines: The Simple Economics of Artificial Intelligence*, illustrated edition edn, Harvard Business Review Press, Boston, Massachusetts.

Agrawal, A. K., Gans, J. S. & Goldfarb, A. (2021), AI Adoption and System-Wide Change, Technical Report w28811, National Bureau of Economic Research.

Alekseeva, L., Azar, J., Gine, M., Samila, S. & Taska, B. (2019), The Demand for AI Skills in the Labor Market, SSRN Scholarly Paper ID 3470610, Social Science Research Network, Rochester, NY.

Alekseeva, L., Azar, J., Gine, M., Samila, S. & Taska, B. (2020), 'The demand for AI skills in the labour market'.

Baldwin, R. (2019), *The Globotics Upheaval: Globalisation, Robotics and the Future of Work*, W&N.

Baldwin, R. & Forslid, R. (2020*a*), Globotics and Development: When Manufacturing is Jobless and Services are Tradable, Working Paper 26731, National Bureau of Economic Research.

Baldwin, R. & Forslid, R. (2020*b*), Globotics and Development: When Manufacturing is Jobless and Services are Tradable, Working Paper 26731, National Bureau of Economic Research.

Benzell, S., Lagarda, G. & Rock, D. (2019), Do Labor Demand Shifts Occur Within Firms or Across Them? Non-Routine-Biased Technological Change, 2000-2016, SSRN Scholarly Paper ID 3427396, Social Science Research Network, Rochester, NY.

Bresnahan, T. F. & Trajtenberg, M. (1995), 'General purpose technologies 'Engines of growth'?', *Journal of Econometrics* **65**(1), 83–108.
**URL:** *http://www.sciencedirect.com/science/article/pii/030440769401598T*

Brynjolfsson, E. & Mitchell, T. (2017), 'What can machine learning do? Workforce implications', *Science* **358**(6370), 1530–1534.

Brynjolfsson, E., Mitchell, T. & Rock, D. (2018), 'What Can Machines Learn, and What Does It Mean for Occupations and the Economy?', *AEA Papers and Proceedings* **108**, 43–47.

Brynjolfsson, E., Rock, D. & Syverson, C. (2017), Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics, Technical Report w24001, National Bureau of Economic Research.

Chiplunkar, G., Kelley, E. & Lane, G. (2020), 'Which Jobs Are Lost during a Lockdown? Evidence from Vacancy Postings in India', *SSRN Electronic Journal* .

Cockburn, I. M., Henderson, R. & Stern, S. (2018), The Impact of Artificial Intelligence on Innovation, Technical Report w24449, National Bureau of Economic Research.

Deming, D. & Kahn, L. B. (2018), 'Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals', p. 33.

Dixon, J., Hong, B. & Wu, L. (2019), 'The Employment Consequences of Robots: Firm-Level Evidence', *SSRN Electronic Journal* .

Felten, E. W., Raj, M. & Seamans, R. (2018), 'A Method to Link Advances in Artificial Intelligence to Occupational Abilities', *AEA Papers and Proceedings* **108**, 54–57.

Felten, E. W., Raj, M. & Seamans, R. (2019), The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization, SSRN Scholarly Paper ID 3368605, Social Science Research Network, Rochester, NY.

Goldfarb, A., Taska, B. & Teodoridis, F. (2020), Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings, SSRN Scholarly Paper ID 3468822, Social Science Research Network, Rochester, NY.

Goos, M. & Manning, A. (2007), 'Lousy and Lovely Jobs: The Rising Polarization of Work in Britain', *The Review of Economics and Statistics* **89**(1), 118–133.

Humlum, A. (2019), 'Robot Adoption and Labor Market Dynamics', p. 79.

Javorcik, B., Kett, B., Stapleton, K. & O'Kane, L. (2020), 'Unravelling Trade Integration: Local Labour Market Effects of the Brexit Vote', p. 40.

Klinger, J., Mateos-Garcia, J. & Stathoulopoulos, K. (2018), 'Deep learning, deep change? Mapping the development of the Artificial Intelligence General Purpose Technology', *arXiv:1808.06355 [cs, econ]* .

Klinger, J., Mateos-Garcia, J. & Stathoulopoulos, K. (2020), 'A narrowing of AI research?', *arXiv:2009.10385 [cs]* .

Korinek, A. & Stiglitz, J. E. (2021), 'Artificial Intelligence, Globalization, and Strategies for Economic Development', p. 41.

Mani, D., Tomar, S., Madan, N. & Bhatia, A. (2020), The Impact of AI on the Indian Labour Market, Technical report, Srini Raju Centre for IT & the Networked Economy, Indian School of Business.

McKinsey Global Institute (2019), AI adoption advances, but foundational barriers remain, Technical report, McKinsey.

NASSCOM (2018), 'Facts & Figures', https://nasscom.in/knowledge-centre/facts-figures.

Oxford English Dictionary (2020), *Oxford English Dictionary*, Oxford University Press.

Perrault, Shoham, Brynjolfsson, Clark, Etchemendy, Grosz, Lyons, Manyika, Mishra & Niebles (2019), The AI Index 2019 Annual Report, Technical report, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA.

Rock, D. (2019), Engineering Value: The Returns to Technological Talent and Investments in Artificial Intelligence, SSRN Scholarly Paper ID 3427412, Social Science Research Network, Rochester, NY.

Seamans, R. & Raj, M. (2018), 'AI, Labor, Productivity and the Need for Firm-Level Data', p. 14.

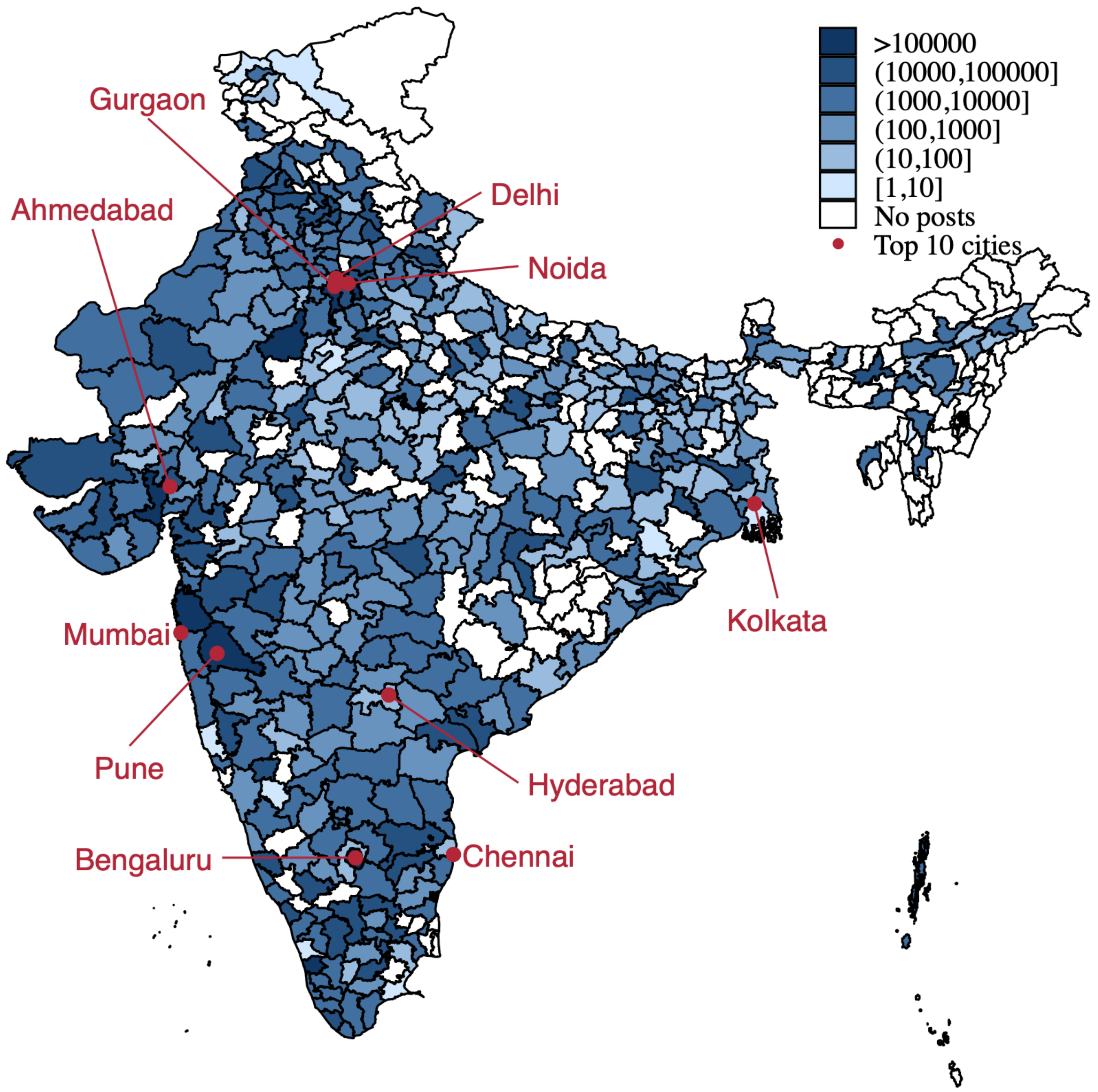
SESEI (2019), Indian ICT Sector Profile Report, Technical report, European Commission.

Stapleton, K. & O'Kane, L. (2020), 'Artificial Intelligence and Service-sector Offshoring', p. 20.

Webb, M. (2020), 'The Impact of Artificial Intelligence on the Labor Market', *SSRN Electronic Journal* .

## Appendix A Additional Figures and Tables

Figure A.1: Total posts by district, 2010-2019




*Notes:* This map shows the distribution of our online vacancy posts across Indian districts. Labels are shown for the ten cities with the largest numbers of posts.

Table A.1: First stage: Impact of AI exposure on establishment AI demand – alternative exposure measures

| | Growth in AI Vacancies | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| AI Exposure | 0.0202*** | 0.0142*** | 0.00629* | -0.0151*** | -0.0102*** | -0.00778** |
| | (5.91) | (4.61) | (1.65) | (-5.71) | (-3.68) | (-2.56) |
| Exposure Measure | Felten et al. | Felten et al. | Felten et al. | SML | SML | SML |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| $R^2$ | .0349 | .0481 | .3774 | .0338 | .0476 | .3775 |
| Observations | 22,251 | 22,251 | 19,383 | 22,251 | 22,251 | 19,383 |

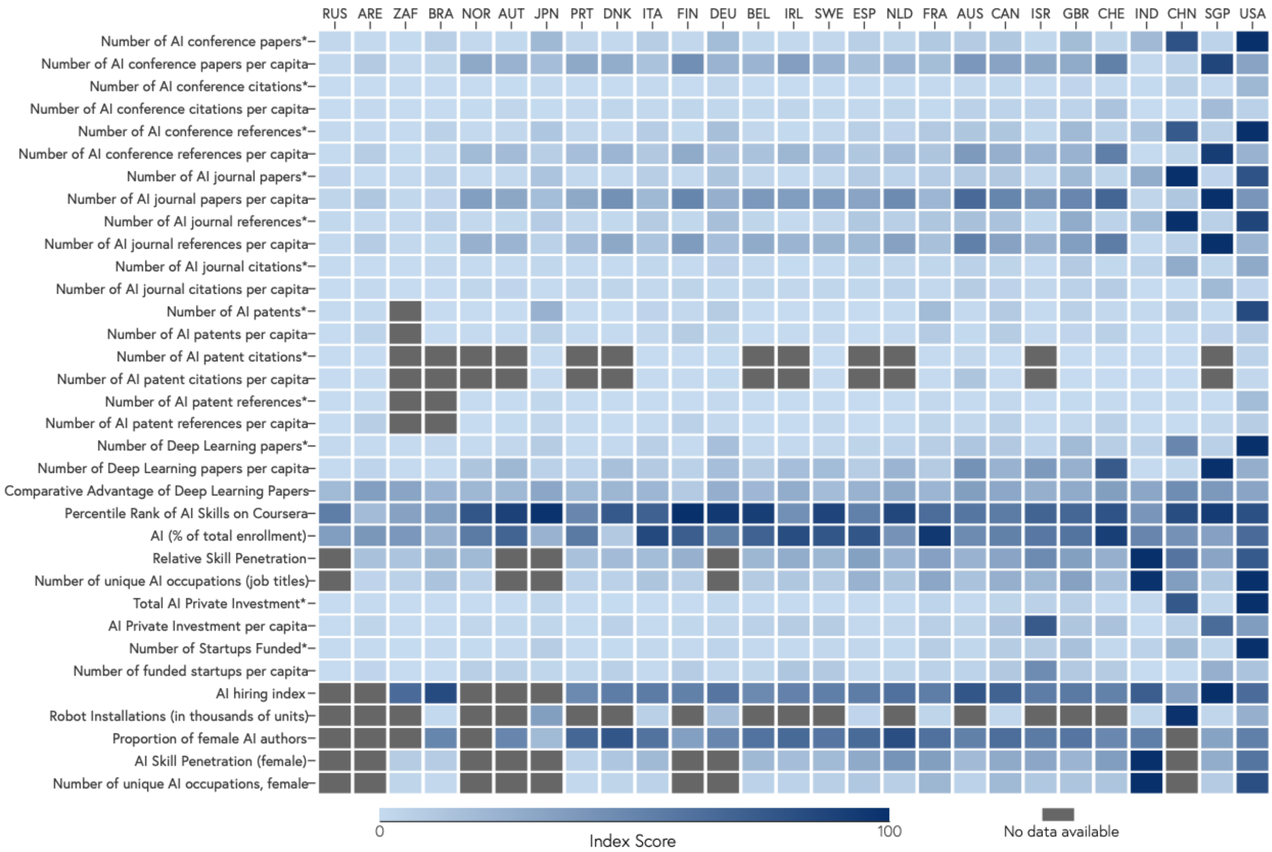
*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. The dependent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. The independent variable is establishment AI exposure, calculated as the standardized average of occupation AI exposure (from either Felten et al. 2018, or Mani et al. 2020 building on Brynjolfsson & Mitchell 2017), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020). Each coefficient therefore represents the proportional impact on AI hiring of a one-standard deviation rise in AI exposure.

Table A.2: Second stage: Impact of AI demand on establishment non-AI vacancies – Felten et al. exposure measure

| | Growth in Non-AI Vacancies | | | Growth in Total Vacancies | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | 1.076 | 0.698 | -1.966 | 1.095 | 0.714 | -1.968 |
| | (1.44) | (0.64) | (-0.79) | (1.47) | (0.66) | (-0.79) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 34.97 | 21.25 | 2.73 | 34.97 | 21.25 | 2.73 |
| Observations | 22,251 | 22,251 | 19,383 | 22,251 | 22,251 | 19,383 |

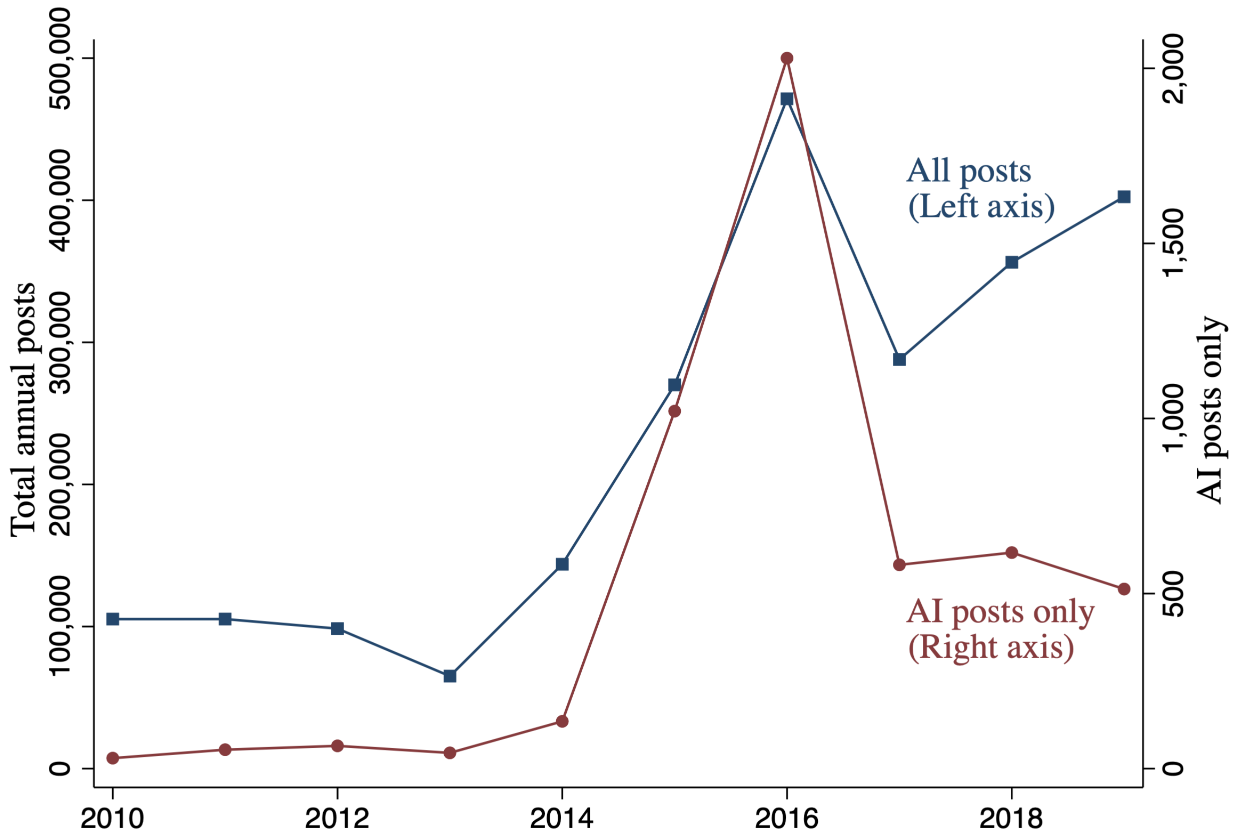
*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Felten et al. 2018), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

Figure A.2: Global AI Vibrancy, from Perrault et al. (2019)



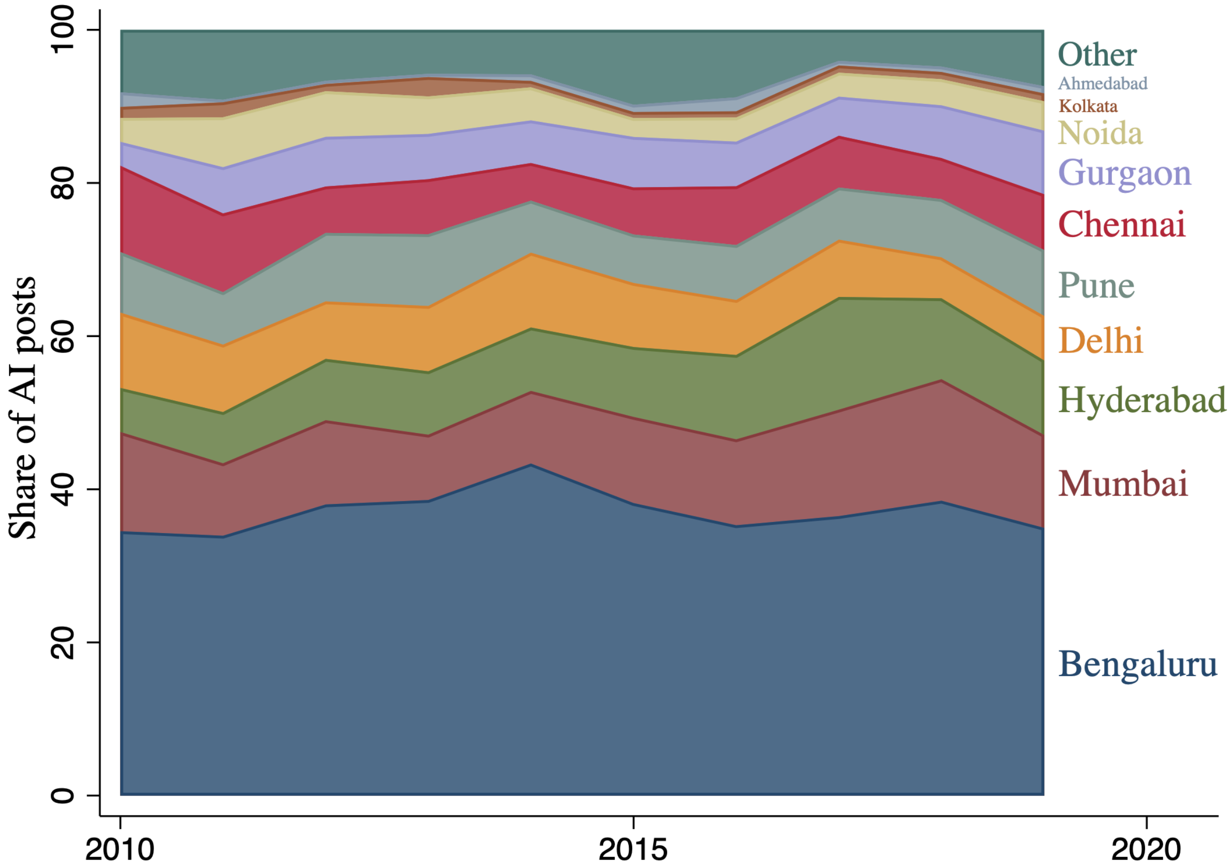

*Notes:* This chart shows relative country scores on a wide range of AI progress metrics. India (fourth from right) scores highly only on skill penetration (the average share of AI skills among all the top 50 skills in each occupation, across all occupations in the country) and number of unique AI occupations (those that have any AI skills in their top 50 skills). These are both calculated using LinkedIn data – which is far less representative in India than in developed countries due to low coverage. Skill penetration is thus likely an overestimate, while the number of AI occupations is largely driven by India's population size.

Figure A.3: AI demand in the Business Process Outsourcing & Call Centre industry




*Notes:* This graph shows the trend in general posts and AI posts in the BPO and call centre industry. The decline in the number of AI posts after 2016 is proportionately larger than the decline in the number of overall posts.

Figure A.4: Cities' shares of AI posts over time




*Notes:* This graph shows the distribution of AI posts across cities over time. Each year reflects the share of all AI vacancies in that year which were in each city. Shares have been remarkably constant. Bangalore's share peaked at just over 40% in 2014, then Mumbai's share in particular has risen subsequently as AI demand increased in finance (see Figure 3.1).

Figure A.5: AI exposure by district – Webb (2020) measure

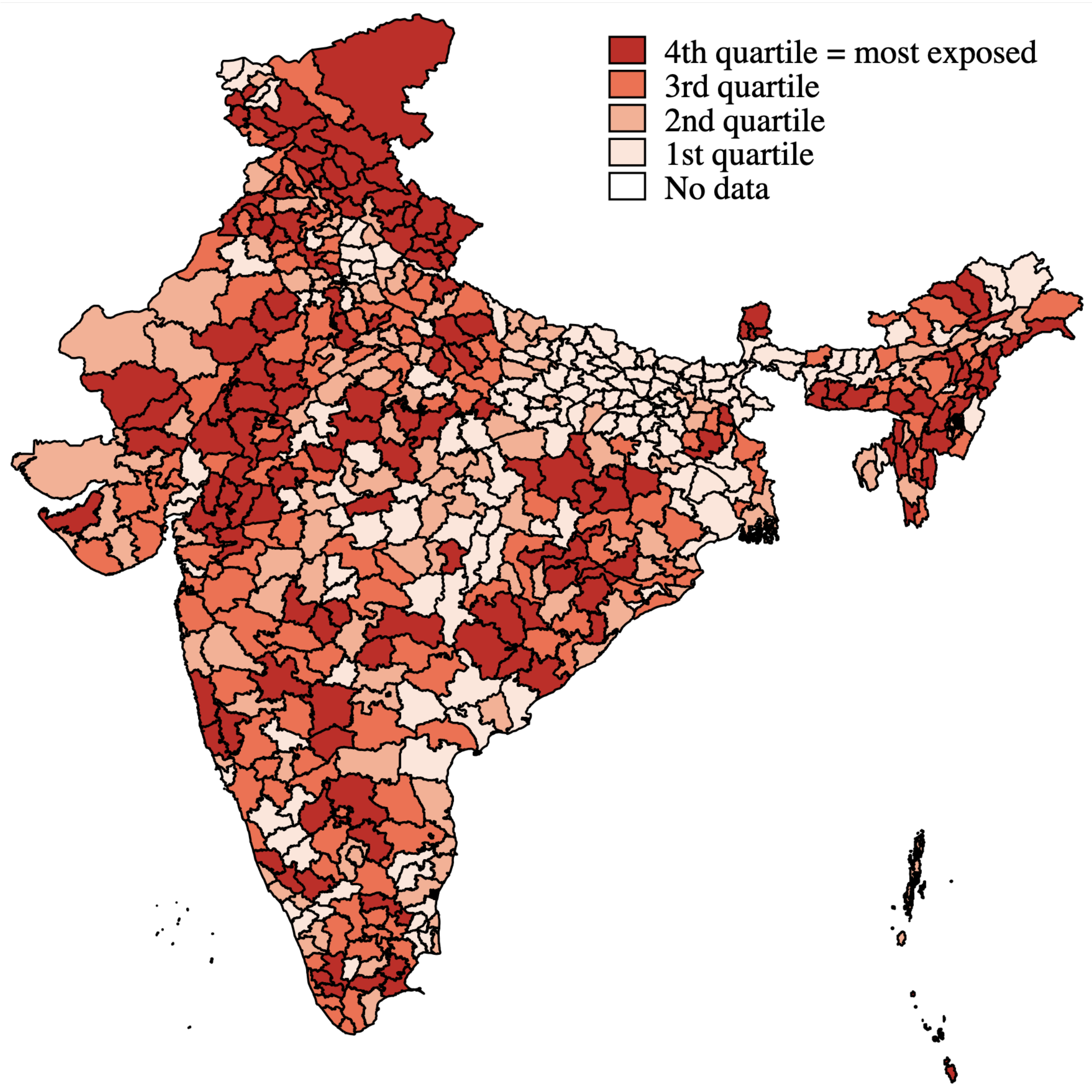
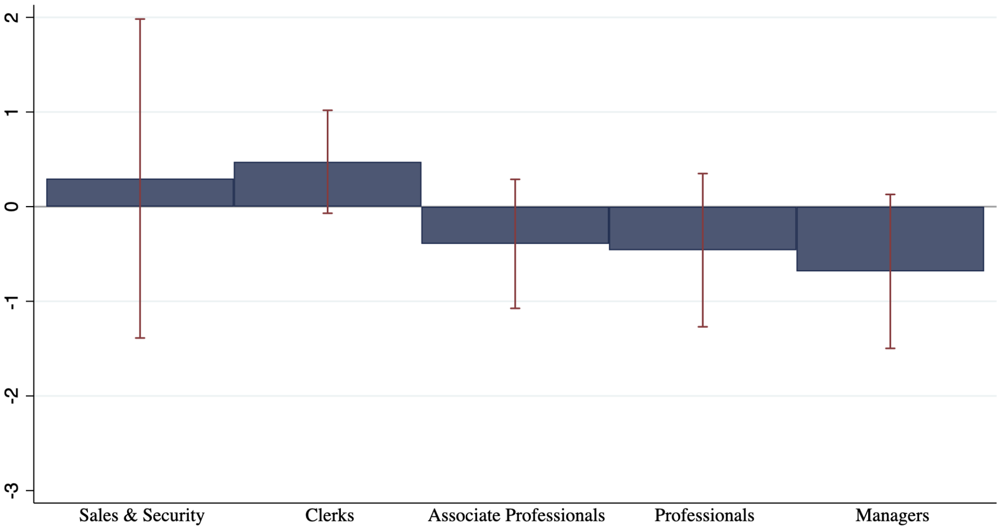
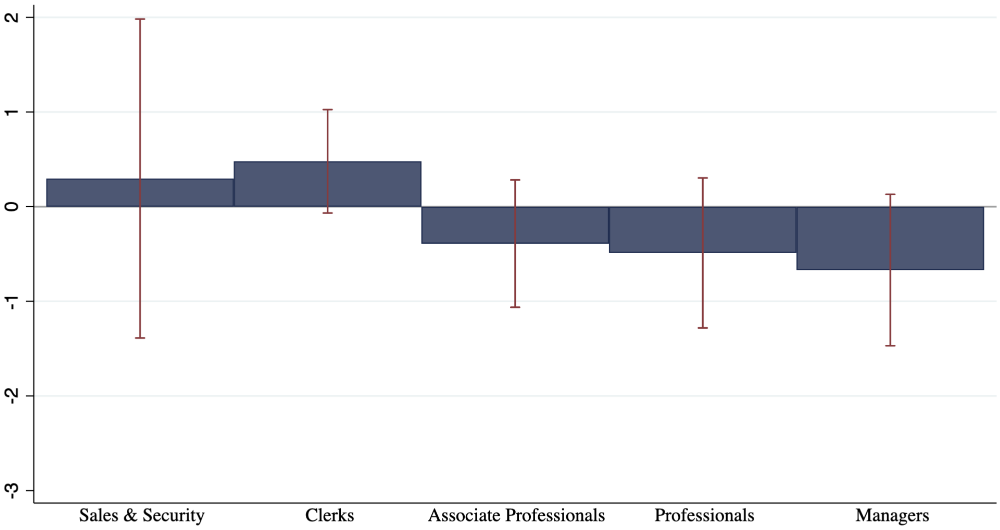


Figure A.6: Impact of AI demand on median wage offers, by broad occupational group

(a) Wage growth in Non-AI posts only



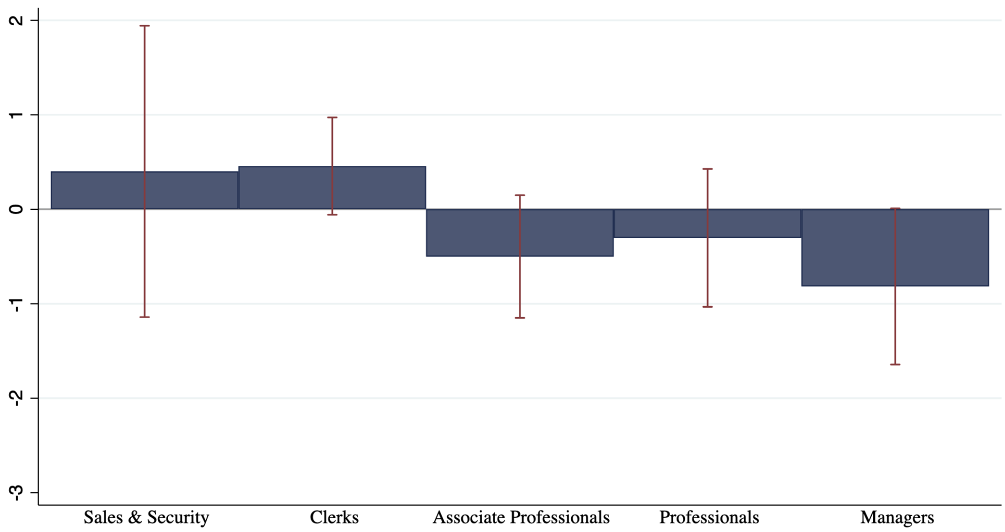

(b) Wage growth in all posts



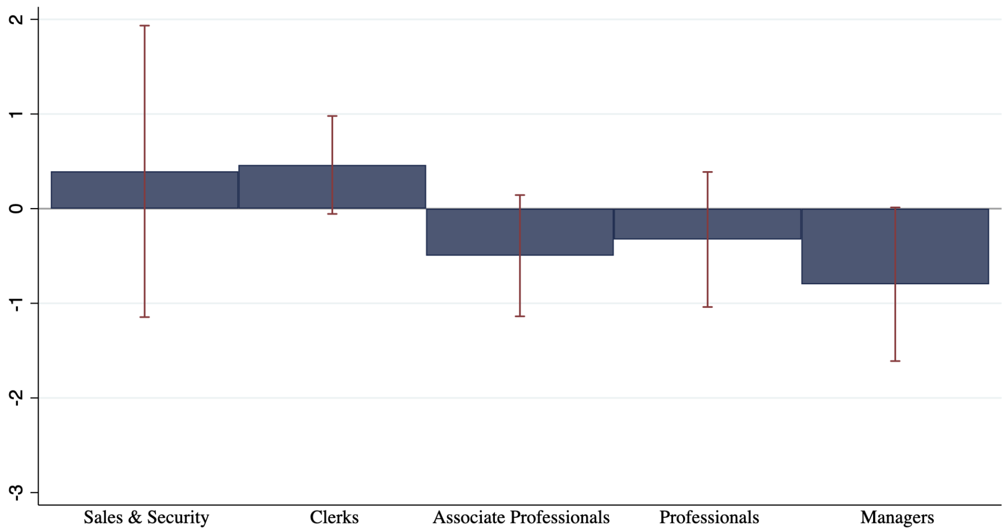

*Notes:* These coefficient plots show the impact of increased establishment AI demand on growth in the establishment median wage, by broad occupational category. Each coefficient in Panel (a) is from a regression of type (2) in Table 5.3, and likewise each coefficient in Panel (b) is from a regression of type (5), except that in each case we construct the outcome variable from establishment wage offers within just one occupational category. Each coefficient thus represents the percentage point impact of a 1% higher growth in establishment AI demand on wage growth in a particular occupational category. As in Table 5.3, AI demand is instrumented with AI exposure, standard errors are clustered at the firm level, and we include region, firm decile and industry fixed effects. Since AI posts make up only a small share of all roles in most establishments, the pattern is very similar across the two graphs.

Figure A.7: Impact of AI demand on mean wage offers, by broad occupational group

(a) Wage growth in Non-AI posts only




(b) Wage growth in all posts




*Notes:* These coefficient plots show the impact of increased establishment AI demand on growth in the establishment mean wage, by broad occupational category. Each coefficient in Panel (a) is from a regression of type (2) in Table 5.3, and likewise each coefficient in Panel (b) is from a regression of type (5), except that in each case we construct the outcome variable from establishment wage offers within just one occupational category. Each coefficient thus represents the percentage point impact of a 1% higher growth in establishment AI demand on wage growth in a particular occupational category. As in Table 5.3, AI demand is instrumented with AI exposure, standard errors are clustered at the firm level, and we include region, firm decile and industry fixed effects. Since AI posts make up only a small share of all roles in most establishments, the pattern is very similar across the two graphs.

Table A.3: Second stage: Impact of AI demand on establishment non-AI wages – Felten et al. exposure measure

| | Growth in Non-AI Median Wage | | | Growth in Overall Median Wage | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -1.108** | -1.512** | -5.856 | -1.133** | -1.567** | -5.797 |
| | (-2.51) | (-2.24) | (-1.48) | (-2.51) | (-2.24) | (-1.49) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 36.02 | 22.15 | 2.567 | 35.05 | 21.22 | 2.607 |
| Observations | 22,064 | 22,064 | 19,217 | 22,071 | 22,071 | 19,223 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Felten et al. 2018), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

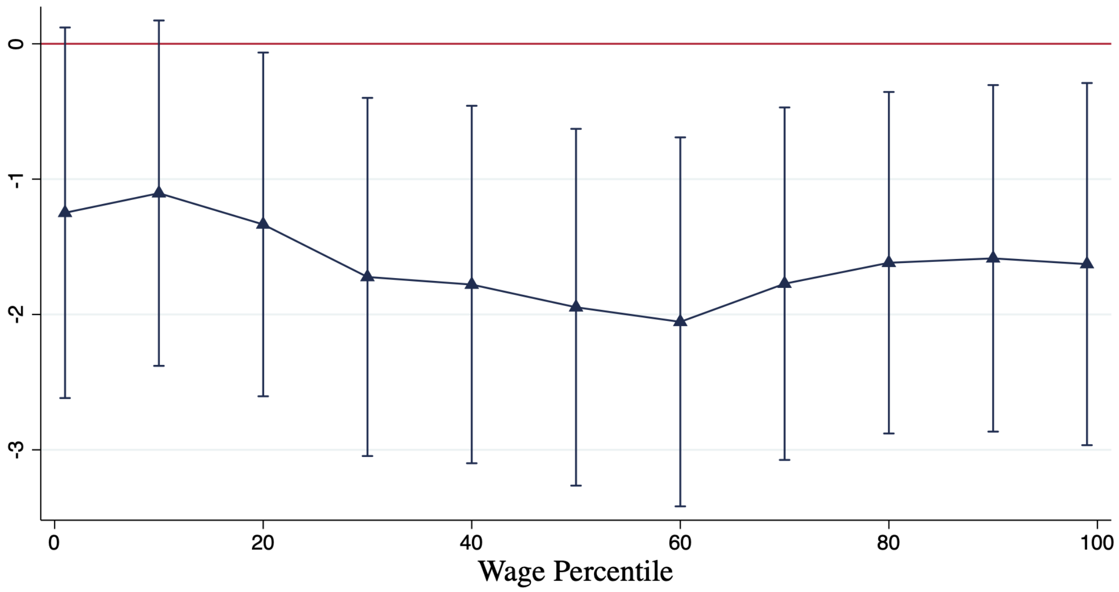
Table A.4: Second stage: Impact of AI demand on establishment non-AI wages, controlling for job profiles – Felten et al. exposure measure

| | Growth in Non-AI Median Wage | | | Growth in Overall Median Wage | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -1.348*** | -1.947*** | -6.437 | -1.374*** | -2.007*** | -6.377 |
| | (-3.17) | (-2.90) | (-1.55) | (-3.14) | (-2.86) | (-1.56) |
| Growth in Experience | 0.826*** | 0.824*** | 0.699*** | 0.826*** | 0.824*** | 0.701*** |
| | (30.71) | (28.73) | (12.98) | (30.61) | (28.46) | (13.25) |
| Growth in High School share | -0.103 | -0.0825 | -0.175* | -0.103 | -0.0812 | -0.176* |
| | (-1.48) | (-0.96) | (-1.77) | (-1.47) | (-0.94) | (-1.79) |
| Growth in Master's share | 0.251*** | 0.257*** | 0.297*** | 0.250*** | 0.256*** | 0.298*** |
| | (7.48) | (7.21) | (3.01) | (7.42) | (7.12) | (3.02) |
| Growth in Doctorate share | 1.847** | 2.399** | 4.245 | 1.876** | 2.461** | 4.191 |
| | (2.17) | (2.10) | (1.24) | (2.17) | (2.09) | (1.24) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 35.06 | 21.5 | 2.597 | 34.09 | 20.57 | 2.639 |
| Observations | 22,064 | 22,064 | 19,217 | 22,071 | 22,071 | 19,223 |

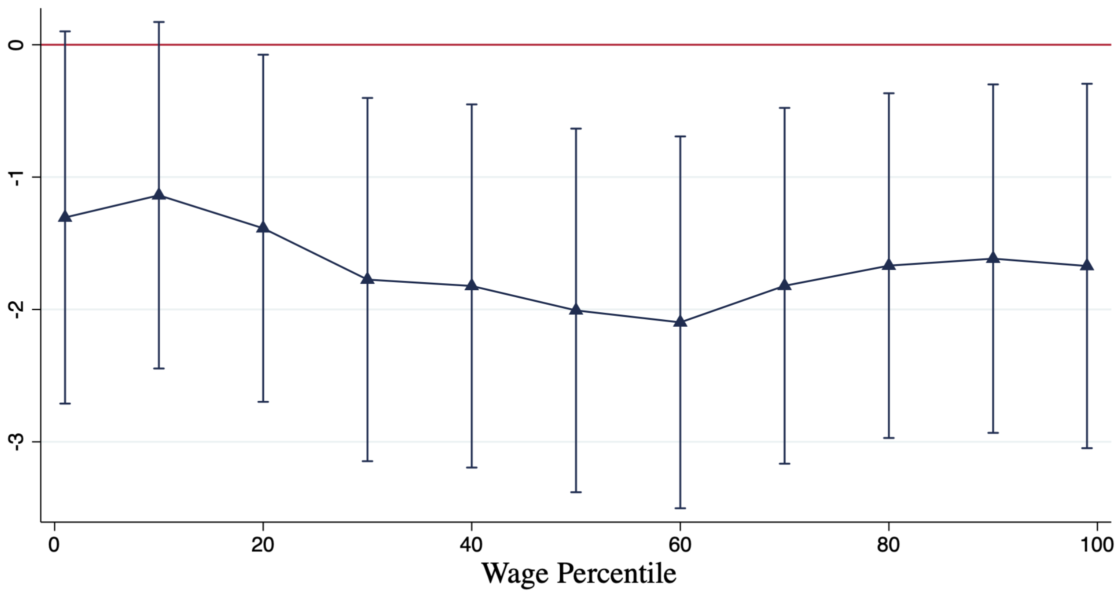
*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Felten et al. 2018), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

Figure A.8: Impact of establishment AI adoption on the wage offer distribution – Felten et al. measure

(a) Non-AI posts only




(b) All posts




*Notes:* These coefficient plots show the impact of establishment AI adoption on the distribution of establishment wage offers. Each coefficient in Panel (a) is from a regression of type (2) in Table 5.4, and likewise each coefficient in Panel (b) is from a regression of type (5). In other words, each coefficient represents the percentage point impact of a 1% increase in establishment AI demand upon a given percentile of the wage distribution. As in Table 5.4, AI demand is instrumented with AI exposure, standard errors are clustered at the firm level, and we include region, firm decile and industry fixed effects. Since AI posts make up only a small share of all roles in most establishments, the pattern is very similar across the two distributions.

Table A.5: Impact of AI demand on establishment non-AI mean wages

| | Growth in Non-AI Mean Wage | | Growth in Overall Mean Wage | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Growth in AI Vacancies | -2.606*** | -1.785*** | -2.531*** | -1.746*** |
| | (-3.59) | (-3.28) | (-3.63) | (-3.32) |
| Controls for Experience & Education | | ✓ | | ✓ |
| *Fixed Effects:* | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | ✓ | ✓ |
| – Industry | ✓ | ✓ | ✓ | ✓ |
| First Stage F-Stat | 26.39 | 26.93 | 27.71 | 28.25 |
| Observations | 22,064 | 22,064 | 22,071 | 22,071 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Webb 2020), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020). Models (2) and (4) also control for changes in establishment job profiles over the period, specifically the mean number of years of experience required and the shares of posts requiring different levels of education.

Table A.6: Second stage: Impact of AI demand on establishment non-AI vacancies, weighted (top 5% winsorized)

| | Growth in Non-AI Vacancies | | | Growth in Total Vacancies | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -1.523** | -0.968* | -5.993 | -1.500** | -0.941* | -5.984 |
| | (-2.42) | (-1.95) | (-1.14) | (-2.39) | (-1.90) | (-1.14) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 18.8 | 16.23 | 1.522 | 18.8 | 16.23 | 1.522 |
| Observations | 22,251 | 22,251 | 19,383 | 22,251 | 22,251 | 19,383 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. Establishments are weighted by baseline number of posts, with the top 5% winsorized. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Webb 2020), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

Table A.7: Second stage: Impact of AI demand on establishment non-AI wages, weighted (top 5% winsorized)

| | Growth in Non-AI Median Wage | | | Growth in Overall Median Wage | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -0.512** | -0.362** | -3.149 | -0.507** | -0.357* | -3.084 |
| | (-2.44) | (-1.97) | (-1.14) | (-2.44) | (-1.96) | (-1.16) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 18.53 | 15.95 | 1.513 | 18.79 | 16.24 | 1.58 |
| Observations | 22,064 | 22,064 | 19,217 | 22,071 | 22,071 | 19,223 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. Establishments are weighted by baseline number of posts, with the top 5% winsorized. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Webb 2020), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

Table A.8: Second stage: Impact of AI demand on establishment non-AI wages, controlling for job profiles, weighted (top 5% winsorized)

| | Growth in Non-AI Median Wage | | | Growth in Overall Median Wage | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -0.331** | -0.248 | -2.841 | -0.329** | -0.244 | -2.784 |
| | (-2.00) | (-1.64) | (-1.19) | (-2.00) | (-1.64) | (-1.21) |
| Growth in Experience | 0.781*** | 0.764*** | 0.730*** | 0.781*** | 0.763*** | 0.729*** |
| | (22.91) | (24.92) | (9.23) | (22.93) | (24.93) | (9.41) |
| Growth in High School share | -0.217** | -0.221** | 0.118 | -0.218** | -0.222** | 0.114 |
| | (-2.24) | (-2.32) | (0.37) | (-2.25) | (-2.34) | (0.37) |
| Growth in Master's share | 0.292*** | 0.312*** | 0.302** | 0.288*** | 0.309*** | 0.306** |
| | (5.93) | (6.55) | (2.22) | (5.84) | (6.47) | (2.24) |
| Growth in Doctorate share | 3.033** | 2.412** | 6.693 | 3.024** | 2.395** | 6.567 |
| | (2.48) | (2.43) | (1.07) | (2.48) | (2.43) | (1.08) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 18.61 | 15.95 | 1.585 | 18.86 | 16.23 | 1.652 |
| Observations | 22,064 | 22,064 | 19,217 | 22,071 | 22,071 | 19,223 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. Establishments are weighted by baseline number of posts, with the top 5% winsorized. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Webb 2020), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

Table A.9: Second stage: Impact of AI demand on establishment non-AI vacancies, weighted (top 0.5% winsorized)

| | Growth in Non-AI Vacancies | | | Growth in Total Vacancies | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -1.091 | -0.826 | -9.747 | -1.071 | -0.801 | -9.740 |
| | (-1.33) | (-1.20) | (-0.41) | (-1.31) | (-1.17) | (-0.41) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 7.28 | 6.69 | .1804 | 7.28 | 6.69 | .1804 |
| Observations | 22,251 | 22,251 | 19,383 | 22,251 | 22,251 | 19,383 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. Establishments are weighted by baseline number of posts, with the top 0.5% winsorized. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Webb 2020), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

Table A.10: Second stage: Impact of AI demand on establishment non-AI wages, weighted (top 0.5% winsorized)

| | Growth in Non-AI Median Wage | | | Growth in Overall Median Wage | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -0.149 | -0.0604 | -4.566 | -0.151 | -0.0602 | -4.429 |
| | (-1.00) | (-0.49) | (-0.42) | (-1.01) | (-0.49) | (-0.43) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 7.236 | 6.646 | .1826 | 7.274 | 6.692 | .1921 |
| Observations | 22,064 | 22,064 | 19,217 | 22,071 | 22,071 | 19,223 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. Establishments are weighted by baseline number of posts, with the top 0.5% winsorized. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Webb 2020), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

Table A.11: Second stage: Impact of AI demand on establishment non-AI wages, controlling for job profiles, weighted (top 0.5% winsorized)

| | Growth in Non-AI Median Wage | | | Growth in Overall Median Wage | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Growth in AI Vacancies | -0.0830 | -0.0252 | -4.174 | -0.0854 | -0.0255 | -4.052 |
| | (-0.65) | (-0.23) | (-0.44) | (-0.67) | (-0.23) | (-0.45) |
| Growth in Experience | 0.720*** | 0.701*** | 0.814** | 0.720*** | 0.701*** | 0.809** |
| | (17.47) | (18.97) | (2.22) | (17.53) | (19.02) | (2.33) |
| Growth in High School share | -0.332** | -0.321** | 0.227 | -0.334** | -0.324** | 0.213 |
| | (-2.47) | (-2.36) | (0.21) | (-2.48) | (-2.37) | (0.21) |
| Growth in Master's share | 0.357*** | 0.381*** | 0.363 | 0.354*** | 0.380*** | 0.361 |
| | (3.47) | (3.65) | (0.86) | (3.44) | (3.65) | (0.88) |
| Growth in Doctorate share | 2.308 | 1.449 | 15.99 | 2.335 | 1.445 | 15.51 |
| | (1.61) | (1.18) | (0.44) | (1.63) | (1.18) | (0.45) |
| *Fixed Effects:* | | | | | | |
| – Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| – Firm Decile | ✓ | ✓ | | ✓ | ✓ | |
| – Industry | | ✓ | | | ✓ | |
| – Firm | | | ✓ | | | ✓ |
| First Stage F-Stat | 7.156 | 6.471 | .1972 | 7.195 | 6.516 | .207 |
| Observations | 22,064 | 22,064 | 19,217 | 22,071 | 22,071 | 19,223 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the firm level. Establishments are weighted by baseline number of posts, with the top 0.5% winsorized. The independent variable is the growth in establishment AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. Likewise the dependent variables are the change in the inverse hyperbolic sine of the respective establishment-level outcomes. Each coefficient therefore represents the percentage point impact upon the outcome variable of a one percent increase in establishment AI hiring. The latter is instrumented by establishment AI exposure. This is calculated as the standardized average of occupation AI exposure (from Webb 2020), over the occupations for which the establishment posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020).

Table A.12: AI demand at the industry-district level

| | AI Vacancies | | Non-AI Vacancies | | Non-AI Wages | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *First stage:* | | | | | | |
| AI Exposure | 0.0126*** | 0.0110** | | | | |
| | (2.82) | (2.14) | | | | |
| *Second stage:* | | | | | | |
| Growth in AI Vacancies | | | -1.038 | -2.375 | 2.264 | -0.166 |
| | | | (-0.68) | (-1.07) | (1.57) | (-0.11) |
| *Fixed Effects:* | | | | | | |
| – District | | ✓ | | ✓ | | ✓ |
| – Industry | | ✓ | | ✓ | | ✓ |
| First Stage F-Stat | | | 7.964 | 4.593 | 7.499 | 4.544 |
| Observations | 4207 | 4165 | 4207 | 4165 | 4204 | 4161 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Standard errors clustered at the district level. The dependent variables are the growth between 2010-12 and 2017-19 in industry-district AI vacancies, non-AI vacancies and non-AI wages, each approximated by the change in the inverse hyperbolic sine. The independent variable for the first stage is industry-district AI exposure, calculated as the standardized average of occupation AI exposure (from Webb 2020), over the occupations for which the industry-district cell posts vacancies in 2010-12, weighted by the number of vacancies posted per occupation, as in Acemoglu et al. (2020). The first two coefficients therefore represent the percentage point impact on AI hiring of a one-standard deviation rise in AI exposure. The independent variable for the second stage is the growth in industry-district AI vacancies between 2010-12 and 2017-19, approximated by the change in the inverse hyperbolic sine. The latter four coefficients therefore represent the percentage point impact upon the outcome variable of a one percent increase in industry-district AI hiring, instrumented by industry-district AI exposure.

Table A.13: AI exposure and wages in alternative datasets

| | Growth in Firm Wage Bill | | Growth in Industry-District Median Wage | | Growth in Industry-District Mean Wage | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| AI Exposure | -0.0866 | -0.0505 | -0.0758*** | -0.0765*** | -0.0798*** | -0.0723*** |
| | (-1.66) | (-0.65) | (-4.12) | (-2.93) | (-4.80) | (-3.10) |
| *Fixed Effects:* | | | | | | |
| – State | | ✓ | | ✓ | | ✓ |
| – District | | | | ✓ | | ✓ |
| – Firm Decile | | ✓ | | | | |
| – Industry | | ✓ | | ✓ | | ✓ |
| $R^2$ | .0092 | .1658 | .0053 | .2372 | .0077 | .2644 |
| Observations | 489 | 460 | 3,451 | 3,297 | 3,451 | 3,297 |

*Notes:* $t$ statistics in parentheses. * p <0.10, ** p <0.05, *** p <0.01. Columns 1-2 use firm-level balance sheet data from services firms in Prowess, and cluster standard errors at the industry level. The independent variable is industry-state AI exposure, calculated using the 2011/12 National Sample Survey. Specifically, it is the standardized average of occupation AI exposure (from Webb 2020), over white-collar services occupations in the industry-state in 2011-12, weighted by the share of workers per occupation. The dependent variable is the growth in the total firm wage bill between 2011-12 and 2017-18, approximated by the change in the inverse hyperbolic sine. Each coefficient therefore represents the proportional impact on total firm wages of a one-standard deviation rise in baseline AI exposure. Columns 3-6 use industry-district-level representative survey data on white-collar services workers, from the National Sample Survey and Periodic Labour Force Survey. Standard errors are clustered at the district level. The independent variable is industry-district AI exposure, calculated using the 2011/12 National Sample Survey. Specifically, it is the standardized average of occupation AI exposure (from Webb 2020), over white-collar services occupations in the industry-district in 2011-12, weighted by the share of workers per occupation. The dependent variable is the growth in the industry-district median or mean wage between 2011-12 and 2017-18, approximated by the change in the inverse hyperbolic sine. Each coefficient therefore represents the proportional impact on average industry-district wages of a one-standard deviation rise in baseline AI exposure.

# Appendix B Data Details

This paper uses four main datasets: vacancy data from India's largest jobs site; balance sheet data from Prowess, which contains longitudinal balance sheet data on all publicly-listed and many large private Indian firms; and nationally representative labour surveys conducted in 2011-2012 (the National Sample Survey) and in 2017-2018 (the Periodic Labour Force Survey). Table B.1 summarises the number of observations across these datasets.

In this Appendix, we provide more details on the composition of the data and the construction of the variables used. We first describe how we classify the occupations, industries and locations in the vacancy data. We then assess the representativeness of the vacancy data by benchmarking it against Prowess and the nationally-representative labour surveys.

Table B.1: Number of observations by data source

| **Online vacancy postings 2010-2019** | #Firms | #Posts |
|---|---|---|
| Agriculture | 13,811 | 463,675 |
| Manufacturing | 57,980 | 2,543,995 |
| Services[1] | 167,969 | 15,481,330 |
| — *Financial* | *17,805* | *1,815,798* |
| — *Information* | *72,057* | *5,834,878* |
| — *Professional* | *38,533* | *834,932* |
| — *Other* | *106,798* | *6,995,722* |
| **Prowess (balance sheets)** | #Firms | #Observations |
| Agriculture | 123 | 590 |
| Manufacturing | 2,276 | 11,257 |
| Services | 3,675 | 16,722 |
| — *Financial* | *1,020* | *4,830* |
| — *Information* | *516* | *2,557* |
| — *Professional* | *199* | *811* |
| — *Other* | *1,940* | *8,524* |
| **Surveys (demographics)** | #Districts | #Households |
| NSS 2012 | 626 | 101,725 |
| PLFS 2018 | 646 | 102,063 |

*Notes:* Some services firms post in multiple sub-sectors, hence the total number of services firms is less than the sum of all firms posting in the sub-sectors.

## B.1 Construction of vacancy dataset

The largest online job postings platform in India scraped and shared 80% of all job postings (randomly sampled). All posts include text data on the job title, industry, role category, location, skills required, salary and experience ranges and educational requirements. We manually mapped 99% of role titles to the 2004 Indian National Classification of Occupations (NCO) at the four-digit level. We also manually mapped all industries to the 2008 Indian National Industrial Classification (NIC) at the two-digit level. We cleaned 95% of city names and added geo-locations, separating out oversees job postings. Using the geolocations, we matched cities to districts, using the 2011 census.

We use publicly-available crosswalks to translate AI exposure measures to the Indian context.

1. Webb: We use a crosswalk to map the 2000 Standard Occupation Classification used by Webb (2020) to the 2004 Indian National Classification of Occupations (NCO), via the 1988 International Standard Classification of Occupations (ISCO), at the four-digit level.
2. Felten: We use a crosswalk to map the 2008 ISCO to the 1988 ISCO, before mapping onto the 2004 NCO.

## B.2 Benchmarking against administrative data

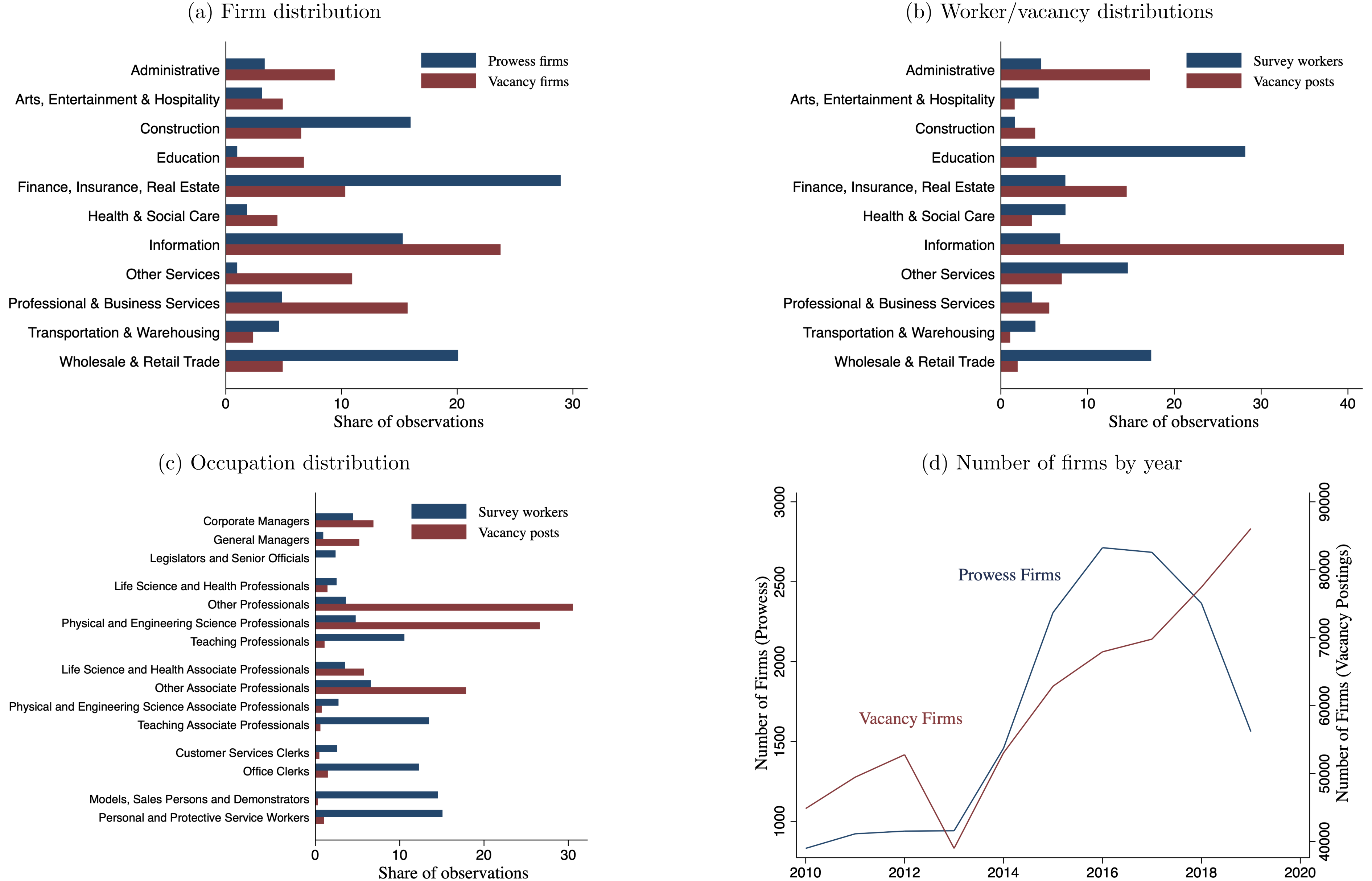
We address the representativeness of our vacancy data in relation to the broader Indian labour market by benchmarking against widely-used administrative datasets. The industry distribution of services firms in the vacancy data and the Prowess dataset are shown in Figure B.1 Panel (a). The distribution of vacancies is shown in Panel (b), alongside the distribution of white-collar services workers in the pooled National Sample Survey (NSS) and Periodic Labour Force Survey (PLFS).[29] The vacancy data has relatively fewer finance, insurance and real estate firms than Prowess, but a greater share in that sector relative to the representative labour surveys. The national surveys also report many more workers in education and transportation, likely because they include public sector workers, whereas the vacancies and Prowess balance sheet data include only private firms. This likely explains in part the large over-representation of the IT sector in the vacancy data, along with IT firms naturally being more comfortable using online tools to advertise vacancies. Panel (c) shows the distribution of services occupations in the vacancy data in contrast to the national surveys. The online posts focus on relatively high-skill white-collar

[29] We define white-collar services workers in the NSS context as salaried workers in divisions 1-5 of the 2004 Indian National Classification of Occupations, i.e. excluding agricultural, fishery, craft, manufacturing, elementary and unclassified workers.

jobs, with fewer roles in lower-skilled jobs, such as shop assistants or security guards, which are more typically filled through referrals and offline hiring.

Our dataset has several advantages over the administrative datasets. The representative sample surveys only took place in 2011-12 and 2017-18, so provide no information on short-term fluctuations or more recent developments in the Indian services sector. Prowess, while useful for studying the largest firms, only contains a limited selection, and did not yet have good coverage for recent years at the time of writing. As illustrated in Figure (d), our vacancy dataset has roughly 30 times the number of firms, and includes many firms in 2018 and 2019. With only balance sheet data, Prowess also offers no clear window on AI exposure or adoption. Similarly, the sample surveys can speak only to AI exposure through the data on occupations, but not adoption. In contrast, we can directly observe AI skills in online job descriptions, the closest category to, for instance, a machine learning engineer in the national survey data is National Occupational Classification code '2132 – Computer Programmers' or '3122 – Computer Assistants'. These broad job classifications alone would be insufficient to identify the use of machine learning, so motivate our move beyond the traditional data sources.

Figure B.1: Comparison of datasets on the Indian services sector




(a) Firm distribution

(b) Worker/vacancy distributions

(c) Occupation distribution

(d) Number of firms by year

*Notes:* These figures compare the composition of our vacancy dataset (red) to that of available administrative datasets (blue). Panel (a) shows the distribution of firms across industries relative to Prowess. Panel (b) compares the distribution of vacancies to that of workers in the NSS & PLFS. Panel (c) shows the distribution of white-collar services occupations relative to NSS and PLFS. Panel (d) compares the number of firms in the vacancy data to that in Prowess.