Working Paper Series

#2026-010

LLM Meets Job Advertisements: Unmasking Skill Premia in the UK

Sidharth Rony

Published 9 June 2026

DOI: https://www.doi.org/10.53330/QGGU3545

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: **info@merit.unu.edu** | website: **http://www.merit.unu.edu**

Boschstraat 24, 6211 AX Maastricht, The Netherlands
Tel: (31) (43) 388 44 00

UNU-MERIT Working Papers
ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology
UNU-MERIT

LLM Meets Job Advertisements: Unmasking Skill Premia in the UK

Sidharth Rony*[1]

[1]UNU-MERIT and Maastricht University.

June 7, 2026

Abstract

Rapid advances in technology and events such as COVID-19 have significantly transformed the modern workplace, potentially altering the skills demanded in jobs. This study examines the evolving demand and posted-wage premia for Information and Communication Technology (ICT), interpersonal, and Artificial Intelligence (AI) skills in the UK labour market. Using a comprehensive dataset of online job advertisements (2016 to 2022), skills are extracted and categorised via GPT-4 zero-shot learning. Cross-sectional log-wage regressions, incorporating occupation and regional fixed effects with three-way Cameron-Gelbach-Miller clustered standard errors, reveal divergent trends in skill compensation. While interpersonal skills are ubiquitously demanded (approximately 90% of listings), they yield no significant posted-wage premium, likely reflecting their near-universal baseline requirement across postings. In contrast, ICT skills, demanded in approximately 55% of postings, carry a posted-wage premium of approximately 7%. AI skills, mentioned in approximately 3% of postings, carry a posted-wage premium of approximately 9% within the ICT-mentioning subsample. These findings document robust associational posted-wage premia for technical competencies amidst recent pandemic-induced and technological labour market shifts.

Keywords: Skills, Wage premium, Machine-assisted mixed methods, Big data, Large Language Model (LLM), COVID-19, AI

JEL Classification: J24, C45, O33

---

*I would like to thank the Department of Economics at Royal Holloway, University of London (RHUL), as well as Prof. Arnaud Chevalier, Prof. Melanie Lührmann, Dr. Stijn Broecke, Prof. Ija Trapeznikova, and other colleagues at RHUL for their valuable guidance and support. I am also grateful to Prof. Andy Dickerson and Prof. Sandra McNally for helpful comments and guidance. I gratefully acknowledge the Urban Big Data Centre for providing access to the Adzuna data. This research was supported by the Economic and Social Research Council (ESRC) through the South East Network for Social Sciences (SeNSS) Doctoral Training Partnership. The views expressed are my own and do not necessarily reflect those of Adzuna, the UBDC, or the ESRC.
Email: sidharthrony@gmail.com

1 Introduction

The labour market is in constant flux, shaped by technological change, economic shocks, and the changing organisation of work. The pace of this transformation has accelerated in recent years, with the continued diffusion of digital technology reshaping the within-occupation skill mix and the wage structure, and the unprecedented disruption of the COVID-19 pandemic compelling firms across industries to reorganise the conduct of work. Demand for software proficiency, digital literacy, and competence in the use of online communication platforms rose sharply, and the disruption raised the value of adaptability and resilience as workers navigated a rapidly changing environment. More recently, AI has emerged as the focal fast-growing sub-area within the broader rising trend in ICT demand, attracting sustained attention in labour-economics research, in industry, and in policy (Autor 2024; Acemoglu et al. 2022). This motivates a separate AI-specific analysis alongside the analyses of ICT and interpersonal skills. Identifying which skills employers value within occupations, characterising how those valuations respond to a shock as large as the pandemic, and assessing whether such changes persist into the subsequent recovery are questions of substantial relevance for policymakers, employers, and workers; they form the focus of this paper.

Despite the salience of these developments, timely and reliable evidence on how employers value particular skills, and on how that valuation responds to large shocks, remains limited. Household and labour-force surveys, administrative earnings records, and the aggregate compensation indicators provided by the national accounts record pay by worker and by occupation rather than by the skill content of the work demanded. These sources are released with substantial lags, are more backward- than forward-looking, and are sensitive to shifts in the composition of workers and hours. They therefore offer limited insight into the value of a particular skill within a given occupation, or into how that value evolves through a disruption as large as the pandemic.

To address this gap, this paper estimates the within-occupation posted-wage premia associated with three families of skills, Information and Communication Technology (ICT), interpersonal, and Artificial Intelligence (AI), using wages and skill content extracted from a near-universe of online job postings in the United Kingdom over the period April 2016 to December 2022. The posting data are drawn from Adzuna, which aggregates vacancies across UK job boards. A distinctive feature of this corpus is that approximately 70% of postings disclose a wage range, rendering UK posted wages an unusually well-populated outcome variable for skill-premium estimation; Adrjan and Lydon (2024) document substantial cross-country variation in wage-disclosure rates, with the UK at the upper end of the distribution.

Skill content is extracted from vacancy text in a two-stage procedure. In the first stage, I follow the ESCO-based pipeline of Kanders and Sleeman (2021) to identify the skills mentioned in each posting, classify each extracted skill into the ICT, interpersonal, or AI category by means of GPT-4 in a zero-shot setting (Karjus 2023), and refine the AI category by imposing a cosine-similarity threshold ($\tau = 0.50$) against the 200-term OECD AI taxonomy of Baruffaldi et al. (2020), using sentence embeddings from the Sentence-BERT model of Reimers and Gurevych (2019). The threshold removes generic terms that constitute common false positives in keyword-based AI measurement.[1] In the second stage, I estimate within-occupation, within-location posted log-wage premia for each skill, with standard errors obtained from the three-way Cameron-Gelbach-Miller procedure clustered on occupation, region, and month (Cameron and Miller 2015).

The paper addresses three questions. First, what posted-wage premium does each of the three skill families carry within occupations and locations? Second, is the within-occupation AI premium sensitive to the AI vocabulary definition, given that keyword-based measures are particularly prone to false positives and that the AI mention rate varies substantially with the threshold imposed? Third, how did each premium evolve before, during, and after the COVID-19 pandemic?

The analysis finds no discernible posted-wage premium associated with interpersonal skills, consistent with their near-universal prevalence in vacancy text. ICT skills carry a premium of approximately 7%, and AI skills (defined at the $\tau = 0.50$ cosine threshold on the ICT-conditional subsample) carry a premium of approximately 9%. The AI premium is robust to the choice of vocabulary: as the cosine-similarity threshold τ rises from 0 to 0.80, the AI mention rate falls from about 13% to under 1%, yet the within-occupation premium stays within a narrow 7% to 11% band across $\tau \in [0.50, 0.75]$, so the premium is not an artefact of the generic terms that inflate prevalence under loose AI definitions. The COVID-19 shock left these premia largely unchanged: the COVID-period interactions are modest for all three skills, and only the ICT $\times$ COVID interaction is economically meaningful.

Three strands of the literature are relevant. The first concerns skills, technological change, and the wage structure. A long-running literature documents the within-occupation shift in skill mix driven by computer adoption and other technologies (Autor et al. 2003; Autor and Dorn 2013; Hanushek and Woessmann 2008), the revised assessment of automation risk once task-level heterogeneity is taken into account (Frey and Osborne 2017; Arntz et al. 2017; Arntz and Böhm 2025), the returns to social and digital skills (Deming 2017; Aghion et al. 2023; Langer and Wiederhold 2023), and the occupation-level AI exposure scoring on which the early AI-labour-market literature has been built (Acemoglu et al. 2022; Felten et al. 2018; Felten et al. 2021; Brynjolfsson et al. 2018; Webb 2019; Eloundou et al. 2024; Brynjolfsson et al. 2023). Occupation-level exposure scores assign a single value to all jobs within an occupation and cannot capture within-occupation variation in skill content; this paper estimates within-occupation premia by exploiting vacancy-level variation in skill mentions, recovering returns that occupation-level scores cannot identify.

The second strand concerns the use of job-vacancy data to study skill demand directly. Hersh-

[1] Section 2 provides the full description of the pipeline.

bein and Kahn (2018) and Modestino et al. (2016) document recession-era skill upgrading and downskilling; Modestino et al. (2023) track skill demand across U.S. online postings; Bessen et al. (2021) examine automation's imprint on posted job content; Dingel and Neiman (2021) identify remote-work-compatible occupations during the pandemic; Gathmann et al. (2024) extract AI-related task changes from German vacancies; and recent contributions extend the approach to digitalisation measurement (Galassi et al. 2024), skill wages from linked vacancy data (Ziegler 2024), joint vacancy- and applicant-side skill dynamics (Bennett et al. 2024), and personality traits from job descriptions (Antonie et al. 2024). The closest UK studies in design are Bone et al. (2025) on AI and green skills (UK Lightcast) and Alekseeva et al. (2021) on within-firm AI premia (U.S. Burning Glass). This paper applies the same vacancy-text approach to the UK Adzuna corpus, a separate near-universe of UK vacancies with substantially higher wage-disclosure coverage than the comparable U.S. corpora, and estimates within-occupation posted-wage premia for ICT, interpersonal, and AI skills jointly rather than for one skill family at a time.

The third strand concerns the methodology for extracting skills from vacancy text using machine learning and natural-language processing. The text-as-data tradition (Gentzkow et al. 2019; Ash and Hansen 2023) has been extended in recent years through large language models for short-text classification (Korinek 2023; Clavié et al. 2023; Chaturvedi et al. 2024). This paper combines GPT-4 zero-shot classification with a cosine-similarity refinement against an external AI taxonomy (Baruffaldi et al. 2020) and reports the AI premium across the full range of the cosine-similarity threshold τ so that the sensitivity of the estimate to vocabulary choice is transparent.

The paper contributes to these literatures on three margins. It develops a two-stage skill-extraction-and-classification procedure (Section 3), combining the ESCO-based pipeline of Kanders and Sleeman (2021) with GPT-4 zero-shot classification (Karjus 2023) and a cosine-similarity refinement against the OECD AI taxonomy of Baruffaldi et al. (2020). It estimates posted-wage premia for ICT, interpersonal, and AI skills jointly within UK occupations on a near-universe of 41 million vacancies over April 2016 to December 2022, extending the single-skill posted-wage analyses of Bone et al. (2025), Alekseeva et al. (2021), Deming (2017), and Aghion et al. (2023) by tracking all three simultaneously. And it documents the evolution of each premium through the COVID-19 shock and the subsequent recovery, using both a three-period interaction specification and a quarter-fixed-effects event study, extending the COVID-period vacancy-text evidence of Dingel and Neiman (2021) and Modestino et al. (2023). The estimates are interpreted throughout as conditional associations between skill content and posted wages rather than as causal returns.

The remainder of the paper is organised as follows. Section 2 presents the data, its preparation, and descriptive statistics, including the cross-source validation of the vacancy dataset against UK official statistics. Section 3 sets out the empirical strategy, and Section 4 reports the main results and robustness checks.

2 Data

The data are the Adzuna UK job-postings corpus, compiled by Adzuna (2025)[2] and spanning April 2016 to December 2022. Each posting records the advertised location, the job title and the corresponding UK ONS SOC code (at 1-, 2-, 3-, and 4-digit levels), the offered wage range, and a free-text description of responsibilities and requirements. The Adzuna corpus captures, on average, 93% of the vacancies identified by the separate ONS vacancy survey (Bassier et al. 2023; ONS 2023), supporting its use as a near-universe of UK job advertisements. The general workflow of data preparation is shown in Figure 2.1 and explained in the subsections that follow.

Data preparation starts with extracting skills using Natural Language Processing (NLP) from job descriptions (Kanders and Sleeman 2021). Subsequently, a comprehensive list is generated, which incorporates all the identified skills. Data cleaning techniques are then applied to eliminate duplicates, resulting in a refined list containing only unique skills. To facilitate further analysis, these skills are then categorised into the broader groups of ICT, interpersonal and AI skills. This classification is achieved via GPT-4 zero-shot learning. Following this categorisation, a mapping process is undertaken to assign the relevant broad skill categories back to their corresponding job advertisements. This step links each job ad with the broader skill sets it requires. Finally, sample restrictions are applied.

[2]The dataset was provided by Urban Big Data Centre under closed license for academic research.

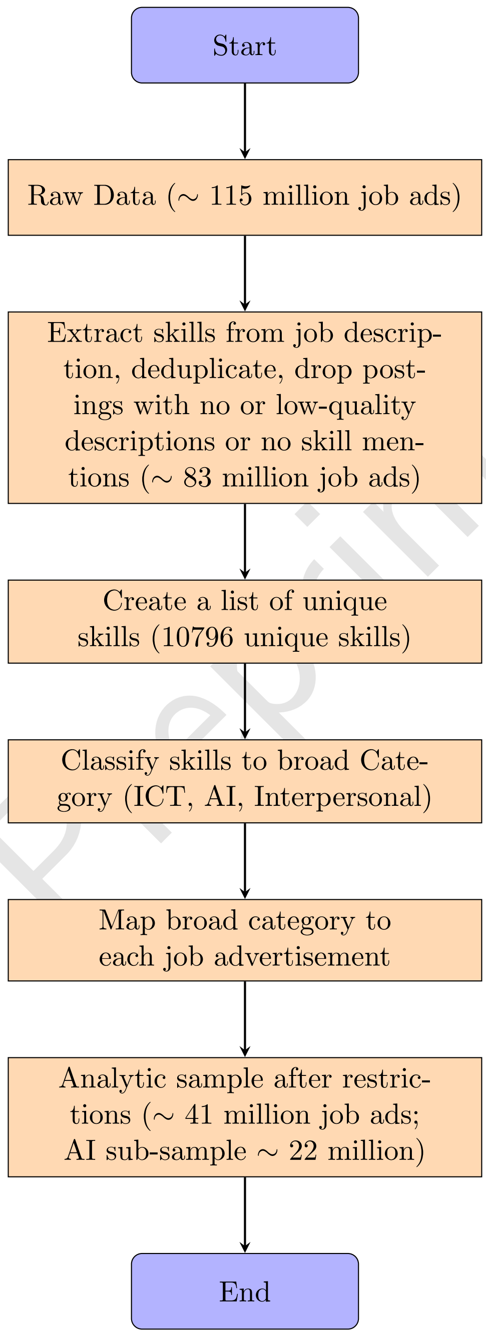

Figure 2.1: General Workflow of Data Preparation

2.1 Skill Extraction and Classification

Skill mentions are extracted from each posting via the ESCO-based surface-form pipeline of Kanders and Sleeman (2021), which maps vacancy phrases to entries in the European Commission's Skill and Competences Organisation (ESCO) framework using a SpaCy-based phrase matcher and uses each ESCO preferred label as the canonical skill identifier. The pipeline yields 10,796 unique skills across approximately 80% of postings; the remaining 20% return no skills, primarily due to poor-quality job descriptions. Figure 2.2 depicts the relative prevalence of the extracted skills.

Figure 2.2: Distribution of Extracted Skills. The size of each term corresponds to its frequency of occurrence in the analysed job-advertisement data.

The 10,796 extracted skills are next assigned to one of three broader categories, ICT, interpersonal, or AI, using GPT-4 in a zero-shot setting. The AI category is further refined by a cosine-similarity threshold ($\tau = 0.50$) against the 200-term OECD AI taxonomy of Baruffaldi et al. (2020) (listed in Appendix A.11), using Sentence-BERT embeddings of Reimers and Gurevych (2019); the refinement removes generic terms and brings the AI mention rate from approximately 13% to approximately 3%. Figure 2.3 summarises the co-occurrence between the three categories at the skill-vocabulary level: of the 10,796 unique skills, 3,315 are classified as interpersonal, 2,447 as ICT, and 302 as AI; 375 are ICT $\cap$ interpersonal, 293 are AI $\cap$ ICT, and 9 are in all three.

The full extraction mechanics, the GPT-4 prompt structure and category definitions, and the trade-offs of the zero-shot classification approach are documented in Appendix A.6, Appendix A.7, and Appendix A.8.

The raw Adzuna corpus over April 2016 to December 2022 contains approximately 83 million postings. After applying wage-disclosure, minimum-wage, and full-time filters (Appendix A.2), the analytic sample used in the ICT and interpersonal regressions of Section 4 comprises ap-

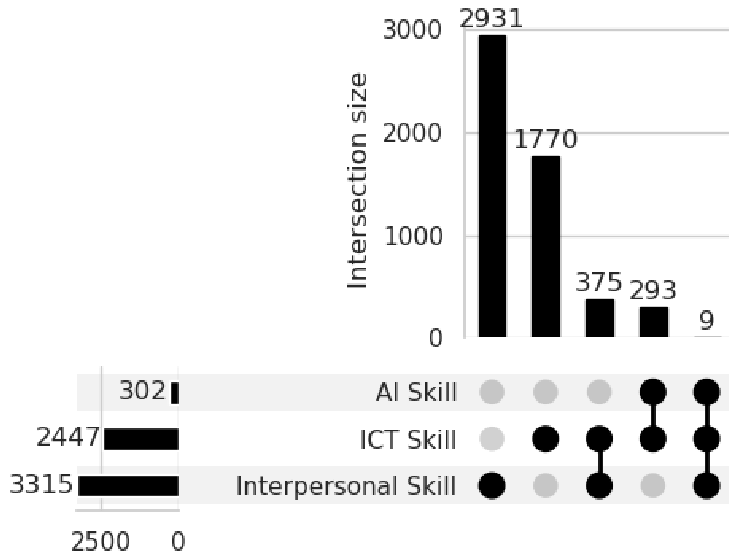

Figure 2.3: Upset plot of the broader skill categories.

proximately 41 million postings; the AI regressions further condition on the vacancy mentioning an ICT skill (because AI skills are, by construction in this corpus, a subset of ICT skills), yielding an AI sub-sample of approximately 22 million. Approximately 70% of UK vacancies in the Adzuna corpus disclose a wage, an unusually high disclosure rate in international perspective (cf. approximately 14% in U.S. online vacancies; Adrjan and Lydon 2024; Batra et al. 2023). Wage disclosure depends modestly on skill content, so the posted-wage premia in Section 4 are interpreted as conditional associations between skill mentions and disclosed posted wages rather than as population-average returns. Sample restrictions, the wage-disclosure selection check, and classifier accuracy are documented in Appendix A.2.

2.2 Descriptive Statistics

The final dataset employed in this analysis incorporates two key wage-related variables: the minimum and maximum wage offered for each job advertisement. Table 2.1 reports their summary statistics on the 83-million-row raw Adzuna corpus, with the AI skill measured at the cosine threshold $\tau = 0.50$ throughout. The mean log maximum wage is 10.02 (SD 1.69) and the mean log minimum wage is 9.86 (SD 1.71); the relatively large standard deviations reflect the wide range of posted wages in the raw corpus, spanning hourly-rate postings at the minimum-wage floor through to senior roles at six-figure annual salaries. The analytic sample of Section 4 restricts wages to between the 2016 annual minimum-wage floor (£13,936) and the 99th percentile of the salary distribution, which compresses this range.[3]

[3] In the further analysis all wages discussed will be the maximum wages offered, unless otherwise specified.

Table 2.1: Summary Statistics on the raw 83-million-row Adzuna corpus, April 2016 to December 2022. Log wage rows use the wage-disclosed subset (approximately 57.8 million postings); skill prevalence and the number-of-skills variable use the full 83M corpus. The AI skill is defined as a posting that the GPT-4 zero-shot classifier labels as AI *and* whose cosine similarity against the OECD AI taxonomy of Baruffaldi et al. (2020) is at least 0.50.

	Observations	Mean	SD
log of maximum wage offered	57,793,930	10.02	1.69
log of minimum wage offered	57,793,930	9.86	1.71
ICT skill	82,825,812	0.53	0.50
Interpersonal skill	82,825,812	0.90	0.30
AI skill ($\tau = 0.50$)	82,825,812	0.03	0.17
Number of skills	82,825,812	6.38	4.71

For each posting, I record whether it demands ICT, interpersonal, or AI skills. Interpersonal mentions are near-universal at approximately 90% of postings; ICT mentions appear in approximately 53%; under the cosine-refined definition at $\tau = 0.50$ (used throughout the analysis), AI mentions appear in approximately 3% of postings, against approximately 13% under the unrefined GPT-4 classification (equivalent to $\tau = 0$). On average, each posting demands 6.4 skills (SD 4.7), reflecting substantial variation in skill breadth across postings.

Each posting carries an occupation code and a geographic identifier. SOC2020 occupation codes are recorded at the 4-digit (unit group), 3-digit (minor group), 2-digit (sub-major group), and 1-digit (major group) levels; coverage is higher at coarser granularities (the SOC-1 major group is recorded for approximately 94% of postings; the SOC-4 unit group for approximately 51%), a trade-off between granularity and coverage that the empirical strategy in Section 3 exploits. Geographic location is recorded at two granularities: the International Territorial Level 1 (ITL1; 12 regions) and the travel-to-work area (TTWA; 228 commuting-defined units).

2.3 Skill Demand

Figure 2.4 plots the monthly count of UK job postings. Postings fell sharply at the onset of the COVID-19 pandemic and recovered through 2021, peaking at the end of the observation window.

The monthly share of postings demanding interpersonal skills remained stable throughout the window (Figure 2.5), at approximately 90% with only minor fluctuation during the COVID period. The ICT share displayed a more dynamic pattern: a sharp decline at the onset of the pandemic was followed by a rise as the pandemic progressed, consistent with accelerated remote-work and digital-tool adoption. The share of postings demanding only interpersonal skills (without ICT or AI) rose during the pandemic and subsequently fell, plausibly reflecting healthcare-sector demand during the acute phase.

Figure 2.6 reports the share of postings in each SOC-1 major group demanding ICT, interper-

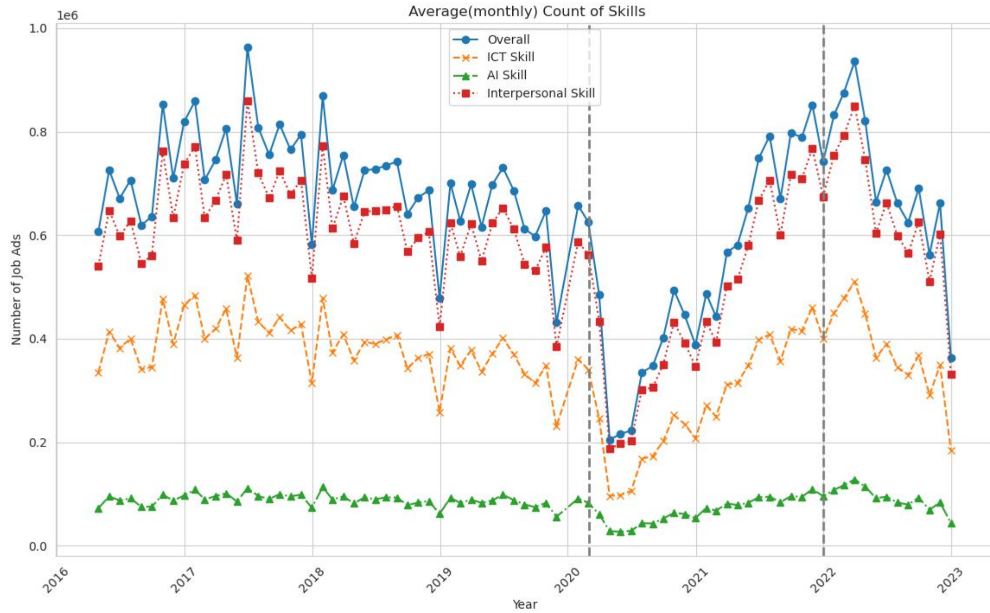

Figure 2.4: Changes in number of job advertisements

sonal, and AI skills on the 83-million-row raw Adzuna corpus. ICT mentions are highest in "Administrative and secretarial occupations" (approximately 67%) and "Associate professional occupations" (approximately 64%); the high prevalence in administrative roles plausibly reflects routine use of office software, data-entry systems, and customer-relationship management tools, while the broad incidence across all groups indicates that foundational ICT use is now expected from managerial through clerical positions. AI mentions, by contrast, are concentrated in higher-skilled occupations: highest in "Professional occupations" (4.9%), followed by "Skilled trades occupations" (2.9%) and "Managers, directors and senior officials" (2.3%), with substantially lower rates in caring, leisure, sales, and elementary occupations (all below 1.5%).

Interpersonal skills are demanded near-universally across all SOC-1 groups: interpersonal mentions appear in between 76% and 99% of postings in every major group, with the highest rates in "Sales and customer service occupations" (99%) and "Caring, leisure and other service occupations" (96%), and the lowest in "Skilled trades occupations" (76%) and "Process, plant and machine operatives" (77%). The relatively small share of director and professional postings without an interpersonal mention plausibly reflects implicit-skill missingness in those job descriptions rather than absent demand, a recognised limitation of vacancy-text data that is taken into account when interpreting the interpersonal premium. The near-universal prevalence is also why the within-occupation interpersonal premium has limited identifying variation in Section 4.1.2.

Skill demand also varies across geography. Figure 2.7 reports the share of postings demanding

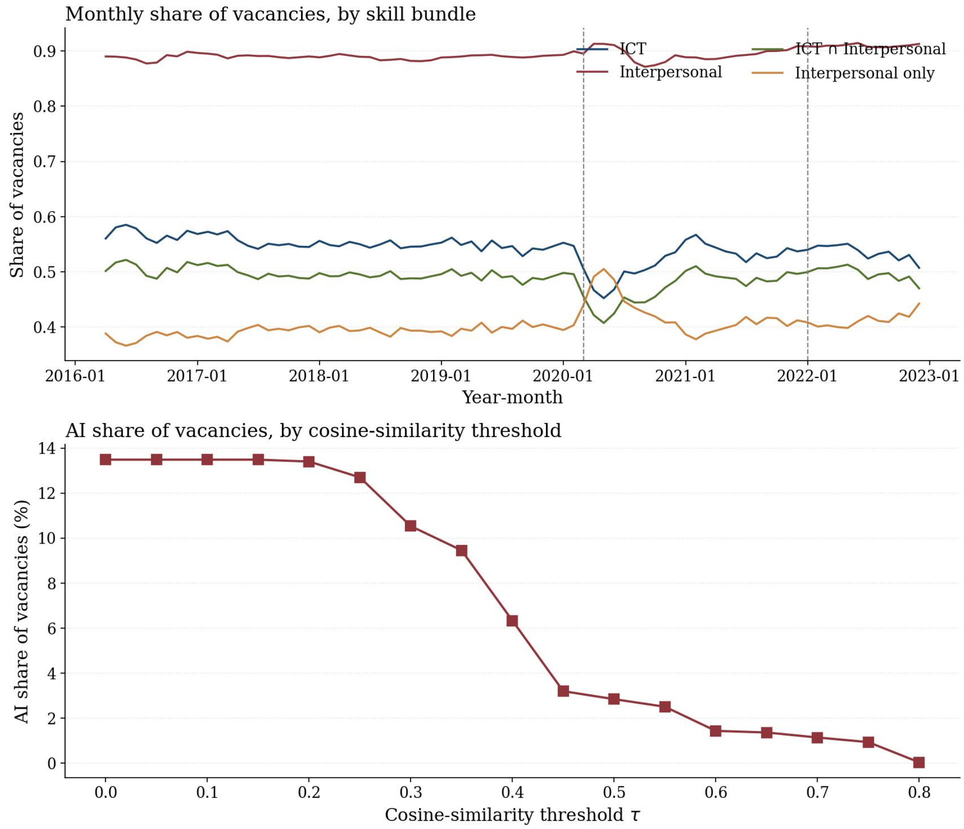

Figure 2.5: Computed on the 83-million-row raw Adzuna corpus (April 2016 to December 2022). *Top panel:* monthly share of UK vacancies demanding each skill bundle (ICT, interpersonal, their intersection, and interpersonal-only) over April 2016 to December 2022. Dashed grey lines mark the COVID-19 onset (March 2020) and the post-COVID cut (January 2022). *Bottom panel:* AI share of vacancies across the cosine-similarity threshold τ used to define the AI skill. At $\tau = 0$ (the unrefined GPT-4 classification) the AI share is 12.79%; at the headline $\tau = 0.50$ used throughout the analysis it falls to 2.85% (cf. Table 2.1).

each skill in each ITL1 region. London has the highest share of AI-mentioning vacancies, at 3.8%; Northern Ireland is second at 3.7% on a substantially smaller posting base (0.8% of the corpus), while the remaining English regions cluster between 2.1% and 2.8%. ICT prevalence is highest in London (60%) and Northern Ireland (63%), and lowest in the North East and Wales (49%). Interpersonal prevalence falls within a narrow band of 89% to 92% across all twelve regions.

Representativeness against ONS. The representativeness of the Adzuna corpus relative to the broader UK online vacancy market is assessed by comparison with the ONS Textkernel online job adverts series over the overlapping January 2017 to December 2022 window (the period covered by both sources). Table A.4 in Appendix A.10 reports the full distributions; the headline patterns are summarised here. Among postings with labelled geography, Adzuna's regional distribution tracks the ONS reference within one percentage point in most regions (London 22.6% in Adzuna vs. 22.7% in ONS; South East 17.7% vs. 15.6%; North West 9.8% vs. 10.0%). Two

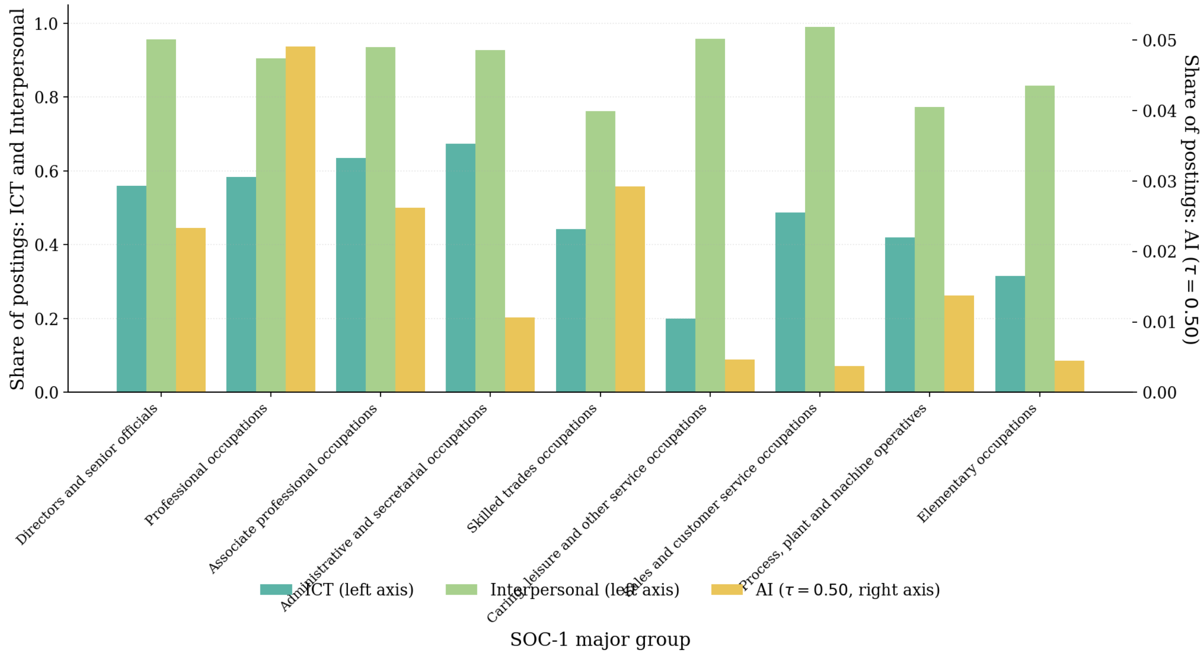

Figure 2.6: Share of postings in each SOC-1 major group demanding ICT, interpersonal, and AI ($\tau = 0.50$) skills, computed on the 83-million-row raw Adzuna corpus (April 2016 to December 2022). The AI skill is defined at the cosine-similarity threshold $\tau = 0.50$ against the OECD AI taxonomy of Baruffaldi et al. (2020).

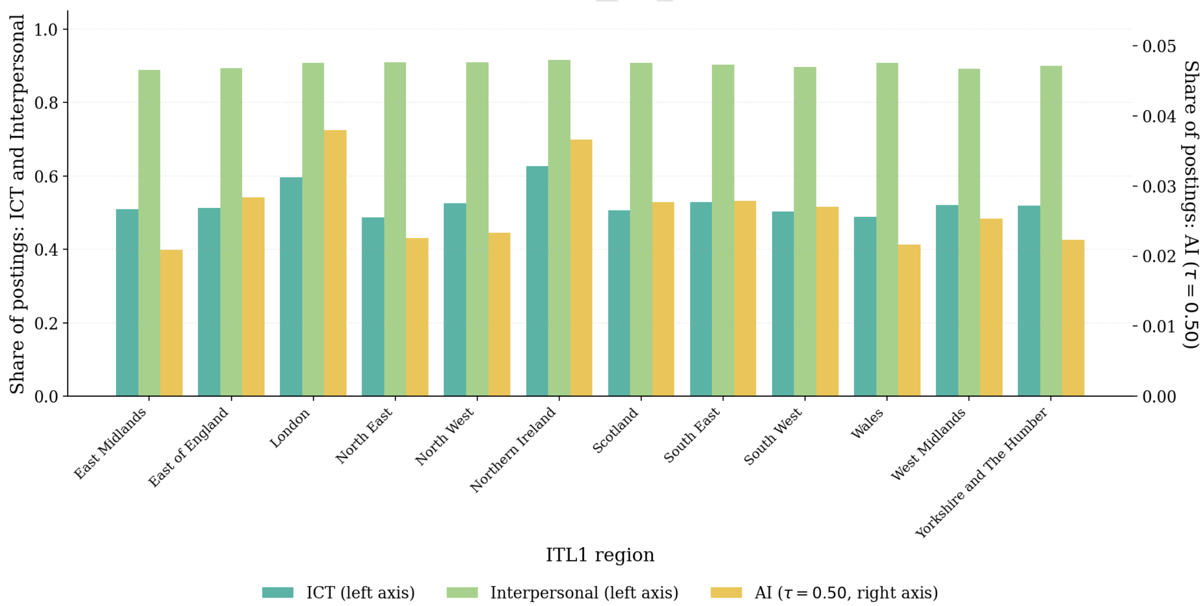

Figure 2.7: Share of postings in each ITL1 region demanding ICT, interpersonal, and AI ($\tau = 0.50$) skills, computed on the 83-million-row raw Adzuna corpus (April 2016 to December 2022). AI skill defined at the cosine-similarity threshold $\tau = 0.50$ against the OECD AI taxonomy.

regions are notably under-represented in Adzuna (Northern Ireland 0.8% vs. 1.5%; Wales 2.5% vs. 3.1%), and the South East is slightly over-represented. 7.7% of the 83-million-row corpus has missing ITL1 and is therefore omitted from this comparison. On occupational coverage, Adzuna over-represents Professional occupations (35.3% of labelled-SOC postings vs. 28.6% in ONS) and Managers, directors and senior officials (9.8% vs. 6.9%), and under-represents Caring, leisure and other service occupations (5.1% vs. 7.8%), Administrative and secretarial occupations (8.1% vs. 10.5%), and Elementary occupations (5.4% vs. 7.7%); this pattern is consistent with the broader finding in the online-vacancy-text literature that online posting platforms over-represent higher-skilled white-collar roles (Hershbein and Kahn 2018; Bone et al. 2025). 5.8% of the corpus has missing SOC2020 major group and is omitted from the SOC-1 comparison.

2.4 Wage

All wage series in this subsection are deflated to April 2016 prices using monthly CPI. Deflation is material for the post-2021 window, in which UK CPI inflation exceeded 5% and the nominal series would otherwise overstate posted-wage growth.

Figure 2.8: Changes in maximum wage offered (inflation adjusted monthly with base of April 2016).

Figure 2.8 shows the time series of real posted wages by skill bundle. The UK pattern stands in apparent contrast to the US evidence of Autor et al. (2024), who document a post-COVID wage compression driven by reallocation of lower-wage workers (e.g. from hospitality) to better-paying jobs in a tight labour market; the UK posted-wage trajectories are instead flat to declining in real terms over the post-2021 window. Whether the difference reflects (i) the posted-wage versus realised-wage distinction (US compression is documented in realised earnings), (ii) the more muted UK labour-market tightness, or (iii) compositional differences in the post-2021 vacancy mix is an open question. The pre-pandemic flat-to-declining trajectory is consistent with the post-2016 deterioration in UK real wages and living standards documented by Costa et al. (2019) and Breinlich et al. (2022). At the onset of the pandemic the offered wage rises temporarily, plausibly reflecting a selection effect (very few vacancies were posted, and those that were tended to be in higher-skilled roles), before falling again to levels below the pre-pandemic baseline. Figure 2.8 also shows a positive association between posted wages and ICT and AI skill mentions, while postings emphasising interpersonal skills track close to the sample average.

Wage disclosure across regions and occupations. The overall wage-disclosure rate on the 83-million-row Adzuna corpus is 69.8%, but the rate is not uniform across postings. Across ITL1 regions, disclosure ranges from a low of 48.5% in Northern Ireland and 64.5% in Scotland to a high of 74.2% in the West Midlands and Yorkshire and The Humber; London is below the GB average at 69.5%. Across SOC-1 major groups, disclosure varies more modestly, ranging from 65.7% in Sales and customer service occupations and 66.3% in Elementary occupations to 74.1% in Administrative and secretarial occupations and 74.0% in Skilled trades occupations. Northern Ireland's relatively low disclosure rate is the dominant geographic driver of the gap between the raw 83-million-row corpus and the 41-million-row analytic sample on which the regressions are estimated: Northern Ireland accounts for 0.8% of the raw corpus, and its further attrition under wage filtering compounds the under-representation of that region in the analytic sample. The full set of disclosure-rate and average-wage cells by region and SOC-1 group is reported in Table A.4 in Appendix A.10.

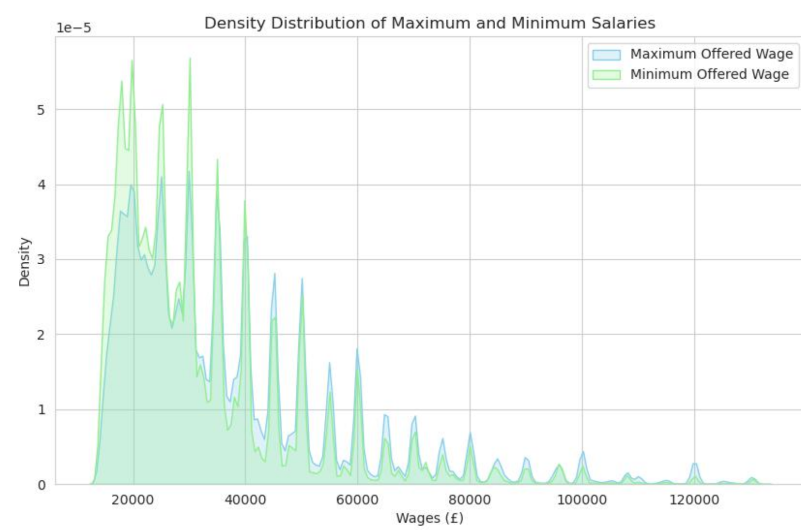

Figure 2.9: Distribution of wages.

Figure 2.9 reports the posted-wage distribution on the analytic sample. Both the minimum and maximum offered wages are concentrated in the £20,000 to £40,000 range, consistent with a large mass of standardised roles in that pay bracket and with the position of the National Living

Wage relative to the offered-wage distribution. The minimum-wage distribution is right-skewed with a distinct peak near £20,000, reflecting a sizeable share of postings at entry-level pay, while the maximum-wage distribution is more dispersed, consistent with a more differentiated market for higher-skilled or specialised roles.

Cross-source validation against LFS realised wages. As an external check on the Adzuna posted-wage measure, posted-wage cell means correlate at $\rho = +0.945$ with realised wages from the UK Labour Force Survey at the SOC-1 $\times$ quarter level (99 cells) and at $\rho = +0.910$ at the SOC-2 $\times$ quarter level (184 cells); the scatter plots and the full methodology are reported in Appendix A.9.

3 Methodology

The analysis estimates two reduced-form specifications. The first treats ICT and interpersonal skills as the joint object and includes their interaction; the second estimates the AI premium on the subsample of vacancies that mention ICT skills. The split is dictated by the construction of the analytic sample: every advertisement classified as demanding AI skills is also classified as demanding ICT skills (the AI $\Rightarrow$ ICT propagation in Section 2), so a unified four-way bundle has an empty ICT$= 0 \times$ AI$= 1$ cell and the AI coefficient would be identified only off the ICT-mentioning subset in any case. Both specifications share the same fixed-effect and clustering structure and the same dependent variable.

Both models utilise a vector of skills, denoted by S_i, to represent the skill set demanded in each job advertisement (i). This vector is a two-dimensional column vector, as shown below:

$$\mathbf{S_i} = \begin{pmatrix} \text{ICT Skill} \\ \text{Interpersonal Skill} \end{pmatrix}$$

The primary specification estimates the joint skill bundle in Equation (1), which includes a set of coefficients capturing the individual effect of each skill (ICT, interpersonal) on the log of the maximum wage offered. These coefficients represent the average wage premium associated with each skill, holding all other factors constant. The model also incorporates interaction terms between pairs of skills ($S_{ij} \times S_{ik}$), which allow us to examine whether the wage effect of a given skill depends on the presence or absence of other skills in the same advertisement. The coefficient on $S_{ij} \times S_{ik}$, for instance, captures how the wage premium of skill j is shaped by the co-occurrence of skill k.

$$ln\_wage_i = \beta_0 + \sum_{k=1}^{2} \sum_{X \subseteq S_i, |X|=k} \beta_X \left(\prod_{S_i \in X} S_i \right) + \text{Time}_i + \text{Location}_i + \text{Occupation}_i + e_i \quad (1)$$

Alongside the joint bundle, I also estimate two single-skill specifications, one for ICT and one for interpersonal, reporting the marginal posted-wage association of each skill without conditioning on the other:

$$ln\_wage_i = \beta_0 + \beta_1 S_i + \text{Time}_i + \text{Location}_i + \text{Occupation}_i + e_i, \tag{2}$$

where S_i is a single binary skill, taking the value ICT_i in the Section 4.1.1 specification and IP_i in the Section 4.1.2 specification. The single-skill estimates appear in Tables 4.1 and 4.2 respectively.

The AI premium uses a separate specification. The regression includes a binary AI variable and is run on the ICT-conditional subsample, since every AI-mentioning posting also mentions ICT skills by construction.

$$ln\_wage_i = \beta_0 + \beta_1 \text{AI\_Skill}_i + \text{Time}_i + \text{Location}_i + \text{Occupation}_i + e_i. \tag{3}$$

$$Time_i = a \cdot m_i + b \cdot m_i^2 + c \cdot m_i^3 + Q2 + Q3 + Q4 \tag{4}$$

The notation $Time_i$ represents a function of time that can capture potential non-linear trends in wages over the study period. This function is specified as a cubic function of a time index (m_i) along with additional dummy variables for quarters ($Q2$, $Q3$, $Q4$) to account for seasonal effects. $Location_i$ is the location of the job advertisement, recorded at the TTWA level.

The term $Occupation_i$ absorbs occupational differences in wage levels, defined at the SOC2020 2-digit sub-major group level from the UK ONS. The 2-digit level groups occupations more coarsely than the 3- or 4-digit codes, trading granularity for computational tractability. Robustness to finer occupational classifications is reported in Appendix B.2, where the premium is estimated across SOC-1 to SOC-4.

$$\begin{aligned}\ln w_i = \beta_0 + \beta_1 S_i + \beta_2 \text{Covid}_t + \beta_3 \text{PostCovid}_t + \\ \beta_4\, S_i \times \text{Covid}_t + \beta_5\, S_i \times \text{PostCovid}_t + \\ \text{Time}_t + \text{Location}_i + \text{Occupation}_i + \varepsilon_i\end{aligned} \tag{5}$$

The second model extends the level-premium specification to allow the skill premium to vary across three macroeconomic regimes. Covid_t marks the pandemic window (March 2020 to December 2021) and PostCovid_t the post-pandemic window (January 2022 onwards); the pre-pandemic period is the omitted baseline. β_2 and β_3 capture the wage differential associated with the pandemic and post-pandemic windows relative to the pre-pandemic baseline, holding skill content fixed. The interaction coefficients β_4 and β_5 measure whether the within-occupation skill premium itself shifts across these regimes. I estimate the model separately for each skill (ICT_i, IP_i, AI_i); each column of Tables 4.5 and 4.6 reports a distinct fit. The ICT and interpersonal columns use the full analytic sample; the AI column uses the ICT-conditional subsample, matching the headline AI specification in Section 4.2.

Standard errors are clustered three-way on occupation (1-digit SOC), location (ITL1 region) and time (month), following Cameron et al. (2011) and the survey of cluster-robust inference in Cameron and Miller (2015), to account for residual correlation within occupation, location, and time cells that single-way clustering would understate.

The cross-sectional log-wage differential associated with a skill, conditional on occupation and location fixed effects, is referred to throughout as a *wage premium*, following standard labour-economics usage. Two qualifications apply. First, the dependent variable is the posted wage on an Adzuna vacancy, not realised earnings on the matched employment outcome; the posted-wage premium need not equal the realised-wage premium if wage disclosure correlates with vacancy characteristics, a point returned to in Section 2. Second, the regressions identify conditional associations, not causal returns to skill. A statement such as "the AI premium is 9.1%" should be read as "vacancies advertising AI skills are associated with posted wages 9.1% higher than otherwise-comparable vacancies, conditional on the included fixed effects".

4 Results

This section reports the estimates from models 1 to 5. Section 4.1 reports the within-occupation premia for ICT and interpersonal skills, together with the joint specification that interacts the two. Section 4.2 reports the AI premium on the ICT-conditional subsample, since AI is by construction a subset of ICT in this corpus. Section 4.3 reports the COVID and post-COVID interactions for each of the three skill groups.

Across all main tables, columns (1) to (4) present a fixed-effect progression (no FE, location only, occupation only, and both). Standard errors are three-way clustered on year-month, SOC-1 major group, and ITL-1 region, following the multi-way clustering procedure of Cameron and Miller (2015). Column (4), which absorbs occupation at the 2-digit SOC level and location at the travel-to-work-area level, is the preferred specification. The coefficients are interpreted as conditional associations between posted wages and advertised skill content rather than causal returns to skill.

4.1 Wage premia for ICT and interpersonal skills

I report the ICT and interpersonal premia in three steps: the single-skill specification for each (Sections 4.1.1 and 4.1.2), and the joint specification with the ICT × interpersonal interaction (Section 4.1.3). The AI premium appears separately in Section 4.2, since AI skills co-occur with ICT skills by construction in the analytic sample.

4.1.1 ICT skill premium

Posted wages are higher in vacancies that mention ICT skills. In the preferred specification (column 4), which absorbs SOC-2 occupation and TTWA location fixed effects and clusters standard errors three-way, the within-occupation ICT premium is 6.9% and is significant at the 5% level. Without occupation fixed effects (column 2) the premium rises to 15.6%, so a substantial part of the unconditional ICT-wage association reflects sorting across occupations rather than the within-occupation return identified by the preferred specification. The estimate is essentially invariant to the choice between ITL1 and TTWA absorption, suggesting that geographic sorting contributes little.

Table 4.1: ICT skill main effect (without interpersonal control)

	log of maximum wage offered			
	(1)	(2)	(3)	(4)
ict_skill	$+0.172^{**}$	$+0.156^{**}$	$+0.077^{**}$	$+0.069^{**}$
	(0.065)	(0.061)	(0.033)	(0.030)
Time Controls	Yes	Yes	Yes	Yes
Occupation FE	No	No	Yes	Yes
Location FE	No	Yes	No	Yes
Observations	41,095,009	41,095,009	41,095,009	41,095,009
r2	0.035	0.098	0.345	0.386
Standard Errors	3-way CGM	3-way CGM	3-way CGM	3-way CGM

Cameron-Gelbach-Miller 3-way clustered SE on (ym, soc2020_major_group, gb_itl1) in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The size of the within-occupation premium is in line with the broader literature on the wage returns to digital and computer-use skills. Krueger (1993) reports a 10% to 15% computer-use premium on U.S. CPS data; DiNardo and Pischke (1997) revisit the question on German tool-use surveys and find similar raw differentials for non-computer tools, cautioning against a literal-return reading of the keyword-based coefficient. Falck et al. (2021) use the OECD PIAAC adult-competencies survey and a broadband-infrastructure instrument to identify a roughly 8% earnings increase per standard deviation of ICT skill. The closest UK study in method is Garcia-

Lazaro et al. (2025), which applies a vacancy-text classifier to UK online postings and reports a 5.8% premium for general digital skills (8.9% for intermediate or advanced digital skills). The 6.9% Adzuna estimate sits inside this band. The premium plausibly reflects the complementarity between digital fluency and the productive tasks demanded within an occupation, in line with the skill-biased-technical-change tradition of Krueger (1993) and Autor et al. (2003) and with the task-level reformulation in Acemoglu and Autor (2011). Throughout, I read the coefficient as a conditional association between the ICT mention and the posted wage rather than as a causal return to skill.

4.1.2 Interpersonal skill premium

Posted wages do not vary significantly with the interpersonal-skill mention within occupations. The column 4 coefficient in Table 4.2 is small (−2.8%) and not statistically distinguishable from zero at conventional levels. Two features of the measurement explain this null. First, interpersonal mentions appear in approximately 90% of vacancies, so the binary measure has very little within-occupation variation to identify a premium from; in such a setting the baseline posted wage already incorporates the employer's expectation of interpersonal proficiency, and the marginal effect of adding the keyword is by construction close to zero. Second, interpersonal expectations are widely taken for granted in posting text and are therefore frequently omitted even when the role requires them, the so-called implicit-skill problem in vacancy-text measurement; an absent mention does not imply that the skill is not demanded.

The null is therefore not evidence that interpersonal skill does not matter for posted wages within an occupation. It is a property of a binary keyword-presence measure applied to a near-universally-mentioned skill. The broader literature using finer measurement strategies finds positive returns: Deming (2017) documents rising returns to social skills in the U.S. labour market over 1980–2012 using a task-based measure; Aghion et al. (2023) reports managerial and social-capital complementarities; and Alekseeva et al. (2021) reports within-firm complementarities between AI demand and managerial skill on U.S. vacancy data. A test of within-occupation returns to interpersonal skill on UK vacancy data would require a taxonomy that separates baseline mentions (*communication*, *teamwork*, *customer service*) from higher-order social-skill content; the present data do not support that distinction. I read the negative point estimate as a conditional association with the binary marker rather than as a causal return to interpersonal capability.

Table 4.2: Interpersonal skill main effect (without ICT control)

	log of maximum wage offered			
	(1)	(2)	(3)	(4)
interpersonal_skill	−0.047	−0.052	−0.030	−0.028
	(0.053)	(0.047)	(0.017)	(0.015)
Time Controls	Yes	Yes	Yes	Yes
Occupation FE	No	No	Yes	Yes
Location FE	No	Yes	No	Yes
Observations	41,095,009	41,095,009	41,095,009	41,095,009
r2	0.003	0.073	0.340	0.382
Standard Errors	3-way CGM	3-way CGM	3-way CGM	3-way CGM

Cameron-Gelbach-Miller 3-way clustered SE on (ym, soc2020_major_group, gb_itl1) in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.1.3 Joint bundle specification

Table 4.3 reports the joint specification on the full sample. Vacancies that mention ICT but not interpersonal skills return a 6.1% premium, closely tracking the 6.9% standalone ICT estimate of Section 4.1.1. Vacancies that mention interpersonal but not ICT, and vacancies that mention both, are not statistically distinguishable from zero in the pooled sample. The structure of the joint specification is informative at the SOC-1 level (Figure 4.1).

Table 4.3: Skill bundle wage premium

	log of maximum wage offered			
	(1)	(2)	(3)	(4)
ict_skill=0 × interpersonal_skill=1	−0.003	−0.009	−0.047	−0.043
	(0.056)	(0.055)	(0.029)	(0.024)
ict_skill=1 × interpersonal_skill=0	$+0.273^{**}$	$+0.253^{**}$	$+0.065^{***}$	$+0.061^{***}$
	(0.082)	(0.081)	(0.015)	(0.014)
ict_skill=1 × interpersonal_skill=1	$+0.159^{*}$	+0.137	+0.033	+0.029
	(0.080)	(0.078)	(0.025)	(0.025)
Time Controls	Yes	Yes	Yes	Yes
Occupation FE	No	No	Yes	Yes
Location FE	No	Yes	No	Yes
Observations	41,095,009	41,095,009	41,095,009	41,095,009
r2	0.037	0.101	0.345	0.387
Standard Errors	3-way CGM	3-way CGM	3-way CGM	3-way CGM

Cameron-Gelbach-Miller 3-way clustered SE on (ym, soc2020_major_group, gb_itl1) in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Robustness checks (Appendix B.1). The appendix reports the bundle specification across the SOC-1 through SOC-3 occupation grid paired with ITL1 or TTWA location absorption, for both maximum and minimum posted wages. Two patterns are consistent. Premia decline monotonically with SOC granularity: finer occupation absorption attenuates the between-occupation sorting that dominates the unconditional ICT premium. The estimates are invariant to ITL1 versus TTWA absorption, indicating that geographic sorting accounts for little of the bundle premium. Minimum-wage estimates track the maximum-wage estimates throughout.

Heterogeneity (Figure 4.1). The within-occupation ICT-only premium is positive and significant in Professional, Skilled trades, Caring/leisure, Process and plant operatives, and Elementary occupations, and is not statistically distinguishable from zero in Directors, Associate professional, Administrative, and Sales. The largest premia are in Caring/leisure and Process and plant operatives, in line with the digital-skill-returns literature cited in Section 4.1.1.

The interpersonal-only premium is significantly negative in Directors, Professional, Associate professional, Administrative, and Sales, and significantly positive in Caring/leisure and Process and plant operatives. The negative readings in the higher-skilled groups are most plausibly an artefact of the implicit-skill problem in vacancy-text measurement rather than a genuine negative return: posting descriptions in director, professional, and associate-professional roles routinely omit the interpersonal mention because the skill is taken as a baseline expectation, so within each of these SOC-1 groups the reference category (vacancies mentioning neither ICT nor interpersonal skills) is disproportionately composed of high-paid postings in which interpersonal skills are demanded implicitly. The interpersonal-only contrast is therefore biased toward negative values by selection rather than by genuine negative returns. The high-skill negative interpersonal estimate should be read as a measurement-error artefact rather than as evidence against the social-skill returns documented by Deming (2017) and Aghion et al. (2023).

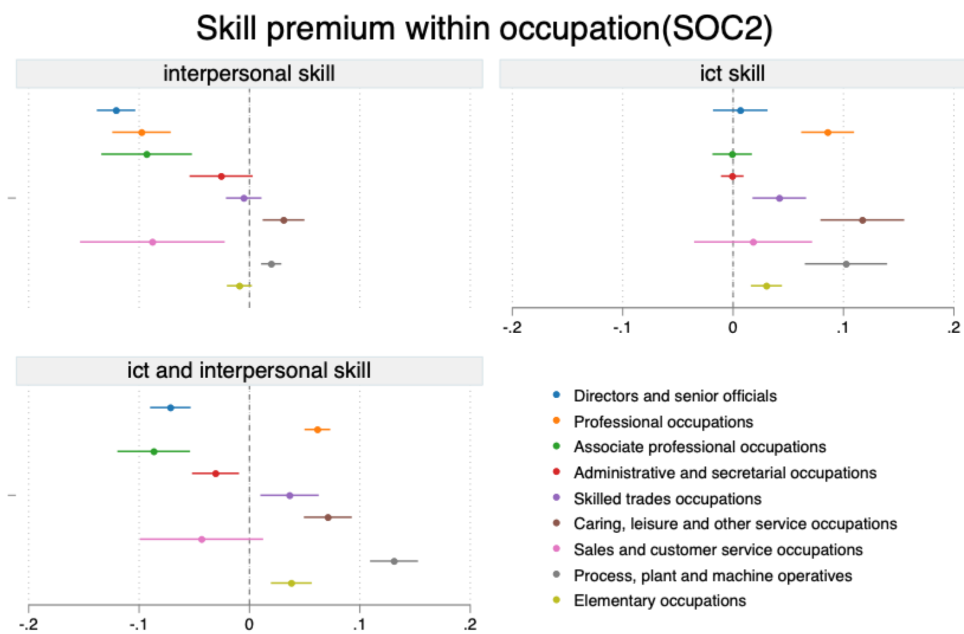

Figure 4.1: Wage premium of the skill bundle across SOC-1 major groups. The corresponding plot by ITL-1 region is reported in Appendix C.1 (Figure C.1).

The joint (ICT and interpersonal) coefficient is significantly negative in Directors, Associate professional, and Administrative, where the same implicit-skill mechanism applies, and significantly positive in Professional, Skilled trades, Caring/leisure, Process and plant operatives, and Elementary. That five of the nine major groups return significantly positive joint coefficients, against a pooled-sample null on the same coefficient, suggests that ICT and interpersonal skills are complementary in some occupational groups rather than uniformly substitutable, in line with the skill-bundle returns documented in Deming (2017) and Aghion et al. (2023). Across ITL-1 regions the pattern varies within a narrow band, so geographic sorting contributes little.

4.2 Wage premium for AI skills (conditional on ICT)

The results provide strong evidence for an AI skill premium in jobs that demand ICT skills (Table 4.4). In the preferred specification (column 4), with SOC-2 occupation and TTWA location fixed effects absorbed and three-way clustered standard errors, the within-occupation AI premium is 9.1%, indicating that even after controlling for occupation and location AI skills command a substantial within-occupation premium. Without occupation fixed effects the premium rises to 23.9% in column 2 and to 26.1% in column 1, so a substantial share of the unconditional AI-wage association reflects sorting across detailed occupations rather than within-occupation returns. The premium is essentially invariant to the choice between ITL1 and TTWA location absorption, suggesting that geographic sorting contributes little.

Table 4.4: AI wage premium

	log of maximum wage offered			
	(1)	(2)	(3)	(4)
ai_skill	$+0.261^{***}$	$+0.239^{***}$	$+0.101^{***}$	$+0.091^{***}$
	(0.043)	(0.042)	(0.010)	(0.010)
Time Controls	Yes	Yes	Yes	Yes
Occupation FE	No	No	Yes	Yes
Location FE	No	Yes	No	Yes
Observations	21,745,023	21,745,023	21,745,023	21,745,023
r2	0.020	0.109	0.355	0.407
Standard Errors	3-way CGM	3-way CGM	3-way CGM	3-way CGM

Cameron-Gelbach-Miller 3-way clustered standard errors in parentheses, clustered on (year-month, SOC-1 major group, ITL-1 region).
AI skill defined at the cosine-similarity threshold $\tau = 0.50$ (see Section 2).
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 4.3 reports the AI premium across a grid of occupation and location fixed-effect choices (SOC-1 through SOC-4 paired with ITL1 and TTWA). The premium is positive and significant in every specification; the magnitude declines monotonically with SOC granularity, indicating that part of the unconditional AI-wage association reflects between-detailed-occupation sorting. The 9.1% AI premium can be placed against several recent estimates. Alekseeva et al. (2021) report a 5% to 11% premium on U.S. Burning Glass vacancies (5% within job title, 11% within firm); Copestake et al. (2022) document a 13% to 17% AI salary premium on Indian job adverts; Engberg et al. (2025) report worker-level evidence from Germany that increased occupational AI exposure raises individual wages; Pouliakas and Santangelo (2025) estimate a positive wage premium for AI programmers relative to comparably educated non-AI programmers using the second European Skills and Jobs Survey across 29 countries; and Bone et al. (2025) estimate a UK Lightcast AI coefficient of $\hat{\beta} \approx 0.23$ to 0.26 on a January 2018 to June 2024 sample with SOC-1, SIC-1, NUTS-1, and year fixed effects (the lower figure adds education and experience controls, which Adzuna does not record at the posting level). As a direct check on the Adzuna corpus, I re-estimate the AI premium on the full analytic sample under Bone et al. (2025)'s specification: SOC-1 occupation, ITL-1 region, and calendar-year fixed effects, without the ICT-conditional restriction. The matched-specification estimates are 21.4% under the OECD taxonomy and 18.5% under the Lightcast vocabulary, both within Bone et al.'s 23% to 30% range (Appendix B.4). The gap between the headline 9.1% and Bone et al.'s estimate is therefore driven by fixed-effect granularity (SOC-2 + TTWA versus SOC-1 + ITL-1) and the ICT-conditional sample restriction, not by data source. Autor (2024) argues, on theoretical and empirical grounds, that AI complements rather than substitutes for middle-skill occupations, consistent with the positive within-occupation AI premium reported here.

A direct vocabulary comparison with Acemoglu et al. (2022) is not well-defined: their AI exposure is a continuous index built from three task-level scores (Felten et al. 2018; Brynjolfsson et al. 2018; Webb 2019) rather than a skill vocabulary. Indirectly, the cosine-refined AI skill at $\tau = 0.50$ excludes the generic terms and retains a technical-AI sub-vocabulary aligned with the Acemoglu et al. and OECD references (voice interaction, motion planning, cluster analysis, autonomous vehicles, image processing, text mining, cognitive computing). The AI premium is stable as τ varies between 0.50 and 0.75 (Figure 4.2), and Appendix B.3 re-estimates the premium across the same range of τ with Bone et al.'s 157-term Lightcast vocabulary in place of the OECD reference: the conclusion that AI carries a meaningful posted-wage premium is invariant to vocabulary choice.

The AI premium decreases as the occupational classification becomes finer. It is 11.9% to 12.2% at SOC-1, 9.1% to 9.3% at SOC-2 (the preferred specification), and 6.2% to 6.4% at SOC-3 and SOC-4. A substantial portion of the AI-wage association at coarser aggregations therefore reflects sorting across detailed occupations rather than within-occupation returns. The estimates are essentially invariant to the choice between ITL1 and TTWA at any given SOC granularity, again pointing to a limited geographic-sorting contribution to the premium.

Robustness checks (Appendix B.2). I re-estimate the AI premium across the full grid of occupation and location fixed-effect choices (SOC-1 to SOC-4 paired with ITL1 and TTWA), for both maximum and minimum wages, and for the AI $\times$ COVID interaction. The within-occupation AI premium is positive and significant in every specification; the magnitude shrinks monotonically with finer SOC granularity, reflecting between-detailed-occupation sorting at coarser aggregations; and the AI $\times$ COVID interaction is small and insignificant throughout. The

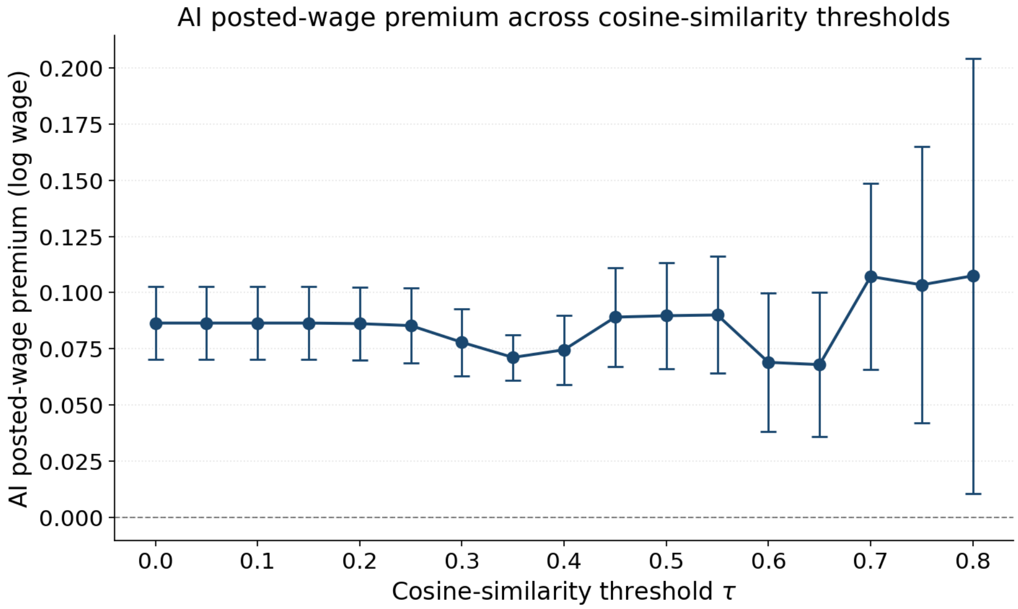

Figure 4.2: AI posted-wage premium ($\hat{\beta}_{\text{AI}}$) with 95% confidence intervals across the cosine-similarity threshold τ used to define the AI skill. Each point is the AI coefficient in the headline log-wage regression with SOC-2 and TTWA fixed effects absorbed, estimated on the ICT-conditional analytic sample. Standard errors are HC1 heteroskedasticity-robust to keep the estimation across the range of τ computationally tractable; at the headline $\tau = 0.50$ the 3-way CGM SE used in Table 4.4 is wider but produces the same qualitative inference. The AI share of vacancies at each τ is plotted separately in Figure 2.5 (bottom panel).

minimum-wage versions track the maximum-wage versions closely.

Heterogeneity (Figure 4.4). The AI premium is positive in every SOC-1 major group, in line with the broader pattern that AI skills are valued across the occupational hierarchy. The magnitudes vary non-monotonically: largest in Process, plant and machine operatives (20.8%), Sales and customer service (18.9%), and Directors and senior officials (14.4%); intermediate in Associate professional, Administrative, and Elementary occupations (each around 11%); and smallest in Professional (7.9%), Caring, leisure and other services (4.7%), and Skilled trades (3.2%). The very large premium in Process, plant and machine operatives, the highest in the SOC-1 grid, was noted as an outlier in the first version of this paper and is plausibly linked to the integration of AI-driven automation into industrial production. The broader pattern that AI's wage effects are concentrated in occupations where AI complements rather than substitutes for occupation-specific human capital is consistent with the task-level reading of Acemoglu et al. (2022) and the complementarity argument of Autor (2024); the present data identify the SOC-1 cross-section of the premium but do not resolve the underlying task-level mechanism.

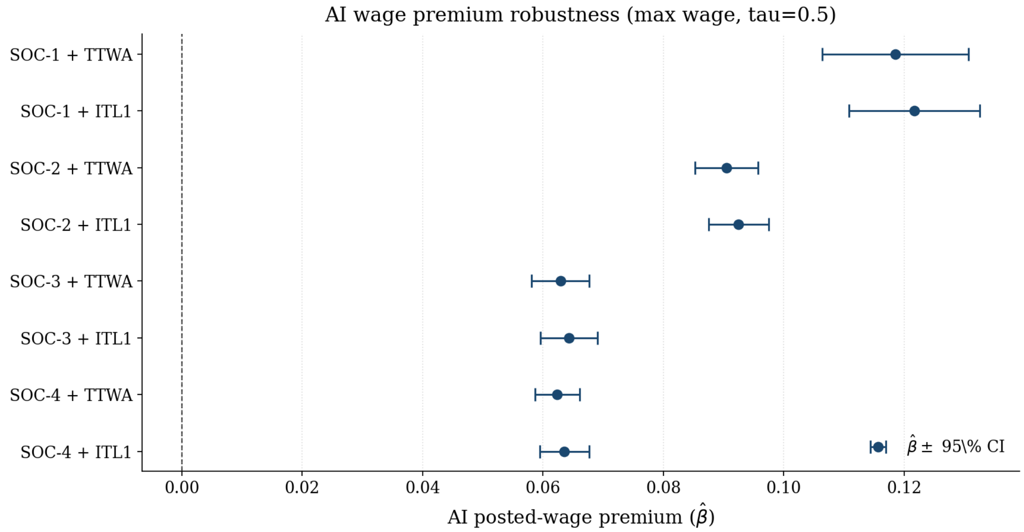

Figure 4.3: AI posted-wage premium ($\hat{\beta}_{\mathrm{AI}}$) and 95% CIs at $\tau = 0.50$ across eight FE combinations: SOC-1/2/3/4 occupation $\times$ TTWA/ITL1 location, estimated on the ICT-conditional analytic sample ($N \approx 22$ million). The headline specification used in the main text is SOC-2 $\times$ TTWA (third row from the top).

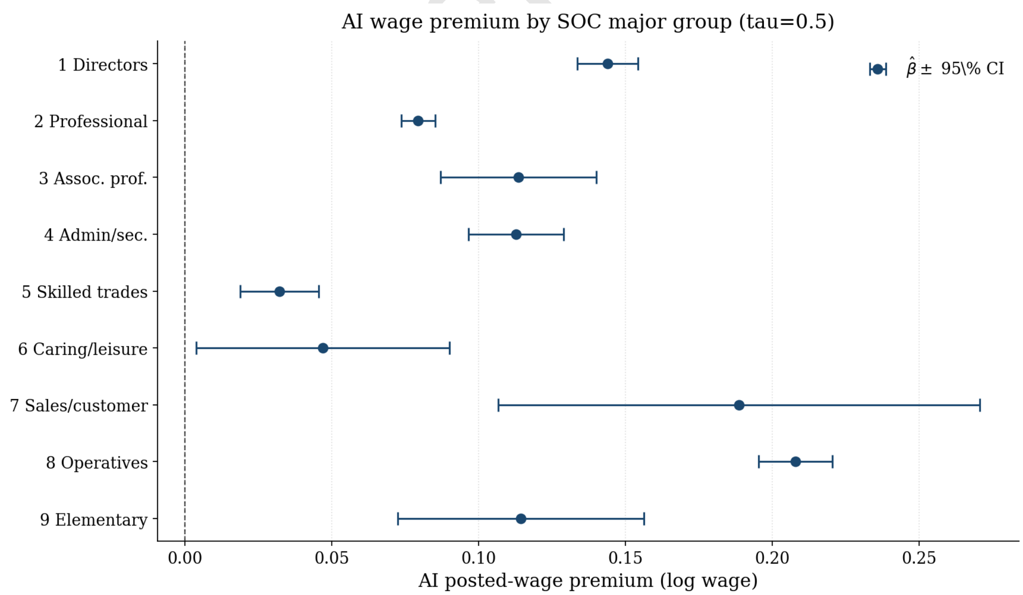

Figure 4.4: AI posted-wage premium $\hat{\beta}_{\mathrm{AI}}$ at $\tau = 0.50$ by SOC-1 major group, estimated on the ICT-conditional analytic sample. The ICT-conditional restriction reflects the construction of the AI skill: every AI-mentioning posting is also an ICT-mentioning posting, so the AI premium is identified off the within-ICT comparison. The corresponding plot by ITL-1 region is reported in Appendix C.1 (Figure C.2).

4.3 COVID-19 and post-COVID interactions with the three skill premia

Model 5 interacts each skill with the COVID and post-COVID period dummies. The ICT and interpersonal interactions are estimated on the full analytic sample (Table 4.5); the AI interactions are estimated on the ICT-conditional subsample (Table 4.6) so as to match the headline AI specification in Section 4.2. The pattern mirrors the level estimates: interpersonal skills do not exhibit a wage premium, while both ICT and AI are associated with positive premia. The wage premium changed only for ICT skills during the COVID-19 period, manifesting as a 1.6 percentage-point increase ($\hat{\beta} = 0.016$, SE $= 0.009$), though the coefficient is not significant at the 5% level under three-way clustering. The AI and interpersonal × COVID interactions are small and insignificant ($\hat{\beta} = -0.008$ and -0.018 respectively), as are all three post-COVID interactions; the interpersonal × post-COVID coefficient ($\hat{\beta} = -0.021$, SE $= 0.009$) is marginally significant at the 10% level, in line with the negative within-occupation interpersonal coefficient in Section 4.1.

A quarter-fixed-effects event study, replacing the cubic time trend with calendar-quarter fixed effects interacted with each skill (28 quarters from 2016Q2 to 2022Q4), is reported in Appendix B.6. The ICT premium is consistently positive and shifts only modestly during COVID, the AI premium is positive throughout with no detectable structural break, and the interpersonal premium hovers near zero throughout, substantiating the headline finding without imposing the cubic functional form.

Robustness checks (Appendix B.5). I re-estimate the COVID interactions across the same fixed-effects grid (SOC-1 to SOC-3 paired with ITL1 or TTWA) and for both maximum and minimum wages. Level premia decline monotonically with finer SOC granularity, are invariant to the choice between ITL1 and TTWA, and the COVID and post-COVID interactions remain small across the grid; only ICT × COVID retains economic significance ($\approx$ 1.5% to 2%). Minimum-wage estimates track the maximum-wage estimates closely.

Compositional decomposition. The aggregate skill-mention shares in Figure 2.5 all fall around the COVID break. An Oaxaca-Blinder shift-share decomposition on the 83-million-row raw corpus separates each change into a within-SOC component (changes in skill-mention density inside each SOC-1 major group, holding the occupational mix fixed at pre-levels) and a between-SOC component (changes in the occupational mix of vacancies, holding within-group skill-mention density fixed). The aggregate ICT-mentioning share falls from 3.5% pre-COVID to 2.3% in the pooled COVID and post-COVID windows (-1.2 pp net), of which -1.1 pp is within-SOC and -0.1 pp between. The aggregate interpersonal share falls from 20.2% to 17.0% (-3.2 pp net), again dominated by the within-SOC component (-3.5 pp, against $+0.4$ pp between). The AI share at $\tau = 0.50$ is small at both points (0.08% pre, 0.07% COVID+) and essentially unchanged. The aggregate decline in raw skill-mention density is therefore a within-occupation phenomenon rather than a compositional reallocation, consistent with a drop in average job-description completeness during the pandemic. The positive within-occupation ICT × COVID

Table 4.5: ICT and Interpersonal wage premia and COVID-19

	log of maximum wage offered	
	(1) ICT	(2) Interpersonal
ict_skill=1	+0.064** (0.027)	
interpersonal_skill=1		-0.021 (0.017)
covid	+0.034* (0.015)	+0.058*** (0.012)
post_covid	+0.053** (0.019)	+0.075*** (0.015)
ict_skill=1 × covid	+0.016 (0.009)	
ict_skill=1 × post_covid	+0.007 (0.009)	
interpersonal_skill=1 × covid		-0.018 (0.012)
interpersonal_skill=1 × post_covid		-0.021* (0.009)
Time Controls	Yes	Yes
Occupation FE	Yes	Yes
Location FE	Yes	Yes
Observations	41,095,009	41,095,009
r2	0.387	0.383
Standard Errors	3-way clustered	3-way clustered

Three-way clustered standard errors on (year-month, SOC-1 major group, ITL-1 region) in parentheses.
Sample: full analytic (N = 41,095,009).
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

interaction in Table 4.5, estimated on the analytic sample conditional on a posting that mentions skills, reflects a different margin: conditional on mentioning skills at all, the within-occupation wage premium for ICT mentions rose modestly during the pandemic.

Heterogeneity across occupations and regions. The SOC-1 occupational and ITL-1 regional cuts of each premium and its COVID interaction are reported in Appendix C.1 (Figures C.3, C.5, and C.7 for the occupational cuts; Figures C.4, C.6, and C.8 for the regional cuts).

ICT. The premium is positive in seven of the nine major groups and largest in directors, pro-

Table 4.6: AI wage premium and COVID-19 (ICT-conditional subsample)

	log of maximum wage offered
	(1) AI
ai_skill=1	+0.093***
	(0.014)
covid	+0.062***
	(0.014)
post_covid	+0.079***
	(0.019)
ai_skill=1 × covid	-0.008
	(0.011)
ai_skill=1 × post_covid	-0.003
	(0.014)
Time Controls	Yes
Occupation FE	Yes
Location FE	Yes
Observations	21,745,023
r2	0.407
Standard Errors	3-way CGM

Cameron-Gelbach-Miller 3-way clustered SE on (year-month, SOC-1 major group, ITL-1 region) in parentheses.
Sample: ICT-conditional subsample (vacancies that mention an ICT skill).
AI skill defined at the cosine-similarity threshold $\tau = 0.50$.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

fessional, and process and plant operatives. It rises mildly during COVID for directors and professional occupations and is essentially flat elsewhere. Across regions, the premium varies within roughly half a percentage point, indicating that geographic sorting contributes little.

Interpersonal. The premium is negative or near zero in every major group, with the largest negative values in director, professional, and associate professional occupations. This is consistent with the interpretation in Section 4.1: in higher-paid occupations, interpersonal expectations are already incorporated into the baseline wage and the binary measure does not differentiate offers. The cross-region pattern is similarly muted.

AI. The premium is positive in every major group, with the largest values in process and plant operatives, sales and customer service, elementary, and directors and senior officials, and the smallest in skilled trades, caring and leisure, and professional occupations. The AI × COVID interaction is small and insignificant in most groups, and cross-region variation is negligible, reinforcing the headline result that AI carries a meaningful within-occupation premium and that the COVID disruption did not produce a detectable structural break in it.

5 Conclusion

The modern workplace has been significantly transformed by rapid technological change and global events such as the COVID-19 pandemic. This study estimates the within-occupation posted-wage premia associated with ICT, interpersonal, and AI skills using a machine-assisted text-classification pipeline applied to UK job advertisements over April 2016 to December 2022. The empirical strategy absorbs occupation and location fixed effects and clusters standard errors three-way on year-month, SOC-1 major group, and ITL1 region.

The study finds that interpersonal skills remain in consistently high demand, with approximately 90% of jobs requiring them. However, no significant posted-wage premium is associated with these skills. This stands in apparent contrast to Deming (2017), who documents that returns to social skills have risen over time in the U.S. labour market. Two related features of the measurement reconcile the null. First, the interpersonal-skill keywords most frequently extracted from UK vacancy text (communication, teamwork, customer service) are baseline competencies that appear nearly uniformly across postings, so the within-occupation variation captured by the binary measure is dominated by generic mentions rather than by the higher-order interpersonal capabilities that Deming's measures emphasise; the widespread necessity of these baseline mentions compresses their measurable posted-wage premium toward zero. Second, the implicit-skill problem in vacancy-text measurement: interpersonal expectations are widely taken for granted in posting text and are therefore frequently omitted even when the role requires them, so the binary mention systematically understates the true incidence of interpersonal-skill demand, and the SOC-1 heterogeneity in Section 4.1 shows this mechanism produces the strongest selection bias precisely in the high-paying occupations where interpersonal proficiency is most implicit.

ICT skills carry a within-occupation posted-wage premium of approximately 7% over the April 2016 to December 2022 window. The raw share of vacancies mentioning ICT fell during the pandemic, and the Oaxaca-Blinder shift-share decomposition in Section 4.3 shows the decline is predominantly within-occupation rather than compositional, consistent with a drop in average job-description completeness during COVID. The within-occupation *wage premium* associated with ICT mentions nonetheless rose modestly during the pandemic, with a +1.6 percentage-point ICT $\times$ COVID interaction in Table 4.5: conditional on a posting that mentions skills, the value placed on ICT capability went up. The pattern fits the broader task-based literature on technology and the wage structure (Acemoglu and Autor 2011; Autor and Dorn 2013), in which the diffusion of computer-using and complementary technologies raises the within-occupation return to technical-skill content, set against a broader backdrop of pandemic-related shifts in education, training, and labour-market outcomes in the UK (Blundell et al. 2021).

AI skills, defined at the cosine-similarity threshold $\tau = 0.50$ against the OECD AI taxonomy of Baruffaldi et al. (2020), are mentioned in approximately 3% of UK postings on the 83 million raw corpus. On the ICT-conditional analytic sample, the within-occupation AI premium is approximately 9% with SOC-2 occupation and TTWA location fixed effects absorbed, rising to 23.9% to 26.1% when occupational fixed effects are not absorbed (Table 4.4). The premium is robust as the cosine-similarity threshold τ varies from 0 to 0.8, over which the retained AI vocabulary tightens substantially (Figure 4.2). The result is in line with Acemoglu et al. (2022), who argue that AI-related jobs yield substantial posted-wage premia due to their specialised nature, and with the UK Lightcast evidence in Bone et al. (2025): when the present specification is matched to Bone et al.'s (SOC-1, ITL-1, and year fixed effects on the full analytic sample), the AI premium on the Adzuna corpus rises to 21.4% under the OECD taxonomy and 18.5% under the Lightcast vocabulary, both within the 23% to 30% range that Bone et al. report on UK Lightcast data. The convergence across two independent UK vacancy corpora, once specifications are matched, is direct external validation of the AI premium estimated here (Appendix B.4). While AI skills do not yet dominate posted-wage demand, they command a substantial within-occupation return, reinforcing the view that specialised technological skills are increasingly valuable.

Four implications follow for the labour-economics literature on within-occupation skill returns. First, the absence of a measurable posted-wage premium on interpersonal-skill mentions does not by itself imply that these skills do not pay off. It is consistent with their being a near-universal prerequisite that no longer differentiates offers and with the implicit-skill problem discussed above; whether higher-order interpersonal capabilities command returns in non-wage dimensions (career progression, job satisfaction, mobility) remains an open empirical question that posted-wage data alone cannot resolve. Second, the within-occupation ICT premium of approximately 7% and the +1.6 percentage-point within-occupation ICT $\times$ COVID interaction in Table 4.5 suggest that the pandemic's primary labour-market signal in the UK posted-wage data is digital intensification of the within-occupation wage return rather than reallocation across occupations. Third, the within-occupation AI premium of approximately 9% on the ICT-conditional subsample, and the convergence with the Bone-matched estimate of 21% on the full sample once specifications are aligned, indicates that AI skills carry an economically meaningful within-occupation return in the UK over the pre-ChatGPT window of this study. Fourth, the Oaxaca-Blinder shift-share decomposition in Section 4.3 establishes that the aggregate decline in skill-mention density during COVID is overwhelmingly a within-occupation contraction rather than a compositional reallocation toward lower-skill occupations: within each SOC-1 major group, the average vacancy mentions fewer ICT and interpersonal skills, while the occupational mix of vacancies barely changes. This is a substantive finding and is, to the author's knowledge, the first such decomposition of UK vacancy-text skill demand around the COVID break.

Several limitations attach to the present estimates and suggest avenues for further work. The dependent variable is the posted wage on an online vacancy rather than realised earnings on the matched employment outcome; the wage-disclosure selection check in Appendix A.2 shows that disclosure varies modestly with skill content, so the estimates are best read as conditional on the joint event of posting and wage disclosure. Because AI vacancies disclose wages less often and tend to occupy higher-paying roles (Appendix A.2), the AI postings missing from the wage-disclosed sample are likely the higher-paid ones, so this selection biases the estimated AI premium downward and the reported coefficients are, if anything, conservative. The GPT-4 zero-shot classifier reflects the model's 2023 training and cannot back-translate to whatever the contemporaneous understanding of "AI" was in 2016 vacancies. This historical-classification concern is mitigated by the cosine-similarity refinement against the OECD AI taxonomy of Baruffaldi et al. (2020), which is an external, predefined reference vocabulary independent of the GPT-4 training corpus; the AI premium remains within a narrow band as the cosine-similarity threshold τ varies from 0 to 0.8 (Figure 4.2), and is essentially unchanged when the OECD reference is replaced by the Lightcast AI/ML vocabulary of Bone et al. (2025) (Appendix B.3). Education and experience information is sparsely populated at the posting level in Adzuna, so including them as controls would substantially shrink the analytic sample to roughly the Bone-equivalent size on which education- and experience-controlled estimates can be obtained, as the Bone-matched specification in Appendix B.4 illustrates. The headline specification preserves a sample much closer to the universe of UK posted vacancies. The sample window ends in December 2022 and does not capture the post-ChatGPT AI hiring surge documented in vacancy data from 2023 onwards. The SIC industry field is sparsely observed in Adzuna and is not used as a fixed effect; industry-specific premia and the industry-composition contribution to aggregate trends cannot be cleanly identified within the present design.

To the author's knowledge, this is the first paper to jointly estimate within-occupation posted-wage premia for ICT, interpersonal, and AI skills on a near-universal UK vacancy panel spanning the COVID-19 shock.

References

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier.

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

Adrjan, P. and Lydon, R. (2024). What do wages in online job postings tell us about wage growth? In *Big Data Applications in Labor Economics, Part A*, volume 52 of *Research in Labor Economics*, pages 71–118. Emerald Publishing Limited.

Adzuna (2025). Economic and social research council. adzuna version 2, 2016-march 2025, 2025 [data collection]. University of Glasgow - Urban Big Data Centre. `https://doi.datacite.org/dois/10.20394%2Ff4vc-yw61`.

Aghion, P., Bergeaud, A., Blundell, R., and Griffith, R. (2023). Social skills and the individual wage growth of less educated workers. Technical report, IZA Discussion Papers.

Alekseeva, L., Azar, J., Giné, M., Samila, S., and Taska, B. (2021). The demand for AI skills in the labor market. *Labour Economics*, 71:102002.

Antonie, L., Gatto, L., Oloumi, S., and Plesca, M. (2024). Personality characteristics demanded by employers: Analysis of job descriptions from university job boards. In *Big Data Applications in Labor Economics, Part A*, volume 52 of *Research in Labor Economics*, pages 213–260. Emerald Publishing Limited.

Arntz, M. and Böhm, M. J. (2025). Technology adoption and labor demand under an aggregate shock. IZA Discussion Paper 17846, IZA Institute of Labor Economics. Documents that firms' COVID-19 exposure had little correlation with technology / AI investment behaviour.

Arntz, M., Gregory, T., and Zierahn, U. (2017). Revisiting the risk of automation. *Economics Letters*, 159:157–160.

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

Autor, D., Levy, F., and Murnane, R. (2003). The skill content of recent technological change: an empirical exploration. *The Quarterly Journal of Economics*, 118:1279–1333.

Autor, D. H. (2024). Applying AI to rebuild middle class jobs. NBER Working Paper 32140, National Bureau of Economic Research.

Autor, D. H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American economic review*, 103(5):1553–1597.

Autor, D. H., Dube, A., and McGrew, A. (2024). The unexpected compression: Competition at work in the low wage labor market. NBER Working Paper 31010, National Bureau of Economic Research.

Baruffaldi, S., van Beuzekom, B., Dernis, H., Harhoff, D., Rao, N., Rosenfeld, D., and Squicciarini, M. (2020). Identifying and measuring developments in artificial intelligence: Making the impossible possible. *OECD Science, Technology and Industry Working Papers*, (2020/05).

Bassier, I., Manning, A., and Petrongolo, B. (2023). Vacancy duration and wages.

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

Bennett, F., Escudero, V., Liepmann, H., and Podjanin, A. (2024). Using online vacancy and job applicants' data to study skills dynamics. In *Big Data Applications in Labor Economics, Part B*, volume 52 of *Research in Labor Economics*, pages 35–99. Emerald Publishing Limited.

Bessen, J., Brynjolfsson, E., Castells-Aparicio, L., and McAfee, A. (2021). Automation, tasks, and skill shifts: Evidence from job openings. *Economics - University of Aarhus, Working Paper (19_05)*.

Blundell, R., Cribb, J., McNally, S., Warwick, R., and Xu, X. (2021). Inequalities in education, skills, and incomes in the UK: The implications of the COVID-19 pandemic. IFS briefing note (part of the IFS deaton review of inequalities), Institute for Fiscal Studies.

Bone, M., Ehlinger, E. G., and Stephany, F. (2025). Skills or degree? The rise of skill-based hiring for AI and green jobs. *Technological Forecasting and Social Change*. Earlier working paper: SSRN 4665577 (2023).

Breinlich, H., Leromain, E., Novy, D., Sampson, T., and Usman, A. (2022). The Brexit vote, inflation and UK living standards. *International Journal of Central Banking*.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2023). Generative AI at work. NBER Working Paper 31161, National Bureau of Economic Research.

Brynjolfsson, E., Mitchell, T., and Rock, D. (2018). What can machines learn, and what does it mean for occupations and industries? NBER Working Paper 24235, National Bureau of Economic Research.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of human resources*, 50(2):317–372.

Chaturvedi, S., Mahajan, K., and Siddique, Z. (2024). Using domain-specific word embeddings to examine the demand for skills. In *Big Data Applications in Labor Economics, Part B*, volume 52 of *Research in Labor Economics*, pages 171–223. Emerald Publishing Limited.

Clavié, B., Ciceu, A., Naylor, F., Soulié, G., and Brightwell, T. (2023). Large language models in the workplace: A case study on prompt engineering for job type classification. In *International Conference on Applications of Natural Language to Information Systems*, pages 3–17. Springer.

Copestake, A., Marczinek, M., Pople, A., and Stapleton, K. (2022). AI and services-led growth: Evidence from Indian job adverts. Imf working paper, International Monetary Fund.

Costa, R., Dhingra, S., and Machin, S. (2019). Trade and worker deskilling. NBER Working Paper 25919, National Bureau of Economic Research.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The quarterly journal of economics*, 132(4):1593–1640.

DiNardo, J. E. and Pischke, J.-S. (1997). The returns to computer use revisited: Have pencils changed the wage structure too? *The Quarterly Journal of Economics*, 112(1):291–303.

Dingel, J. I. and Neiman, B. (2021). Remote work during the covid-19 pandemic: Evidence from online job postings. *Proceedings of the National Academy of Sciences*, 118(33):e2110003118.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, 384(6702):1306–1308.

Engberg, E., Koch, M., Lodefalk, M., and Schroeder, S. (2025). Artificial intelligence, tasks, skills, and wages: Worker-level evidence from Germany. *Research Policy*, 54(8):105285.

Falck, O., Heimisch-Roecker, A., and Wiederhold, S. (2021). Returns to ICT skills. *Research Policy*, 50(7):104267.

Felten, E. W., Raj, M., and 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., and Seamans, R. (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal*, 42(12):2195–2217.

Frey, C. B. and Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change*, 114:254–280.

Galassi, G., Bellatin, A., and Chu, V. (2024). Letting job postings talk: Recent trends in digitalization. In *Big Data Applications in Labor Economics, Part B*, volume 52 of *Research in Labor Economics*, pages 1–33. Emerald Publishing Limited.

Garcia-Lazaro, A., Mendez-Astudillo, J., Lattanzio, S., Larkin, C., and Newnes, L. (2025). The digital skill premium: Evidence from job vacancy data. *Economics Letters*, 250:112154.

Gathmann, C., Grimm, F., and Winkler, E. (2024). AI, task changes in jobs, and worker reallocation. IZA Discussion Paper 17554, IZA Institute of Labor Economics.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–574.

Hanushek, E. A. and Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3):607–668.

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

Kanders, K. and Sleeman, C. (2021). Open Jobs Observatory: Extracting skills from online job adverts. `https://www.nesta.org.uk/project-updates/skills-extraction-ojo/`.

Karjus, A. (2023). Machine-assisted mixed methods: augmenting humanities and social sciences with artificial intelligence. *arXiv preprint arXiv:2309.14379*.

Korinek, A. (2023). Generative ai for economic research: Use cases and implications for economists. *Journal of Economic Literature*, 61(4):1281–1317.

Krueger, A. B. (1993). How computers have changed the wage structure: Evidence from microdata, 1984–1989. *The Quarterly Journal of Economics*, 108(1):33–60.

Langella, M. and Manning, A. (2022). Residential mobility and unemployment in the uk. *Labour Economics*.

Langer, C. and Wiederhold, S. (2023). The value of early-career skills. CESifo Working Paper 10288, CESifo. Documents rising returns to digital and social skills since the 1990s in German apprenticeship-acquired skills.

Modestino, A., Shoag, D., and Ballance, J. (2016). Downskilling: changes in employer skill requirements over the business cycle. *Labour Economics*, 41:333–347.

Modestino, A. S., Burke, M. A., Sadighi, S., Sederberg, R. B., Stern, C. A., and Taska, B. (2023). No longer qualified? changes in the supply and demand for skills within occupations. Working Paper WP 23-15, Federal Reserve Bank of Boston.

ONS (2023). Using Adzuna data to derive an indicator of weekly vacancies: Experimental Statistics. `https://tinyurl.com/pft5hkzm`. [Online; accessed 08-March-2024].

Pouliakas, K. and Santangelo, G. (2025). Are artificial intelligence (AI) skills a reward or a gamble? deconstructing the AI wage premium in Europe. IZA Discussion Paper 17607, Institute of Labor Economics (IZA).

Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Webb, M. (2019). The impact of artificial intelligence on the labor market. *Available at SSRN 3482150*.

Ziegler, L. (2024). Skill demand and wages: Evidence from linked vacancy data. In *Big Data Applications in Labor Economics, Part B*, volume 52 of *Research in Labor Economics*, pages 101–130. Emerald Publishing Limited.

A Data Appendix

A.1 Hardware and parallel-extraction configuration

Skill extraction from a 115-million-observation job-description corpus strains computational resources, and a parallel processing framework was implemented. The extraction used a server equipped with 64 cores. Read and write operations were executed serially to ensure data integrity, while the core skill-extraction tasks were performed in parallel across 60 of the 64 cores. With this configuration the full extraction completed in approximately five wall-clock days. I also implemented the Nesta phrase-matcher and surface-form filtering in a streaming fashion to keep peak memory within the server's available RAM. The matching-algorithm details and the rule-based-pattern specifics are documented in Kanders and Sleeman (2021), to which the interested reader is referred for the full algorithmic description.

A.2 Data Preparation

Skill extraction. Skill mentions are extracted from each posting via the ESCO-based surface-form pipeline of Kanders and Sleeman (2021), which yields 10,796 unique skills across approximately 80% of postings; the remaining 20% return no skills (deduplication, descriptions with text but no ESCO-codable mention, and postings with no usable description). The surface-form mechanics of the pipeline are documented in Appendix A.6, the parallel-extraction hardware configuration in Appendix A.1, and the relative prevalence of the extracted skills in Figure 2.2.

Skill classification. The 10,796 extracted skills are classified into ICT, interpersonal, or AI categories using GPT-4 in a zero-shot setting (carried out 17 to 19 November 2023), in place of the manual and PCA-style dimensionality-reduction methods reviewed in Langella and Manning (2022) which scale poorly to short-text skill vocabularies. Operating at the preferred-label level (10,796 inputs) rather than the job-title level used by Clavié et al. (2023) keeps classification cost low while preserving the language model's general world knowledge. The category definitions, written to fix the conceptual scope ex ante, are:

- ICT skill: *"An ICT skill involves the ability to use Information and Communication Technology tools and applications effectively, such as computers, software applications, and digital communication platforms."*
- Interpersonal skill: *"An interpersonal skill is the ability to effectively communicate, interact, and work well with other people."*
- AI skill: *"For a human, an AI skill involves the ability to design, interact with, or implement artificial intelligence technologies and applications."*

The ESCO skill framework provided the basis for the initial definitions. Initial AI-classification accuracy in the 60% to 70% range, driven by GPT-4 treating "tasks executable by AI" as AI skills regardless of whether human design or implementation was required, was resolved by the human-anchored AI definition above. The GPT-4 API call structure (prompt template, skill-by-skill input, binary category output) is documented in Appendix A.7; the trade-offs of the zero-shot approach are discussed in Appendix A.8.

Classifier accuracy. The classifier is validated on a random sample of 100 skills (Table A.1): interpersonal 96%, ICT 92%, AI 84%.

Table A.1: Skills and Accuracy

Skill	Accuracy
ICT skill	92%
Interpersonal skill	96%
AI skill	84%

Measurement error. Two sources of measurement error affect the headline regressions and motivate the downstream refinements. First, classifier noise generates classical attenuation bias: with reliability r in the binary skill variable, OLS attenuates by roughly $1-r$, so the AI premium (84% accuracy) is biased toward zero by approximately 16%, with smaller bias for ICT (92%) and interpersonal (96%). As a partial remedy, the AI skill is further restricted to matches with cosine similarity above a threshold τ (Section 4, robustness across the range of τ). A related concern, specific to AI: skills classified by 2023-trained GPT-4 may not have been recognised as AI in 2016 vacancies, plausibly producing a modest upward bias on the historical AI mention rate; for consistency, these terms are labelled as AI throughout. Second, the extraction is run on the full posting rather than zoned to the requirements section; mentions originating in firm-description, employer-branding, or benefits segments contribute a weakly upward incidence bias. Zoning via a transformer-based section classifier is left to future work.

Co-occurrence and AI refinement. Figure 2.3 (upset plot) and Figure A.1 (cosine-similarity matrix at the broader-category level) summarise the overlap between the three categories at the skill-vocabulary level. Of the 10,796 unique skills, 3,315 are classified as interpersonal, 2,447 as ICT, and 302 as AI; 375 are ICT $\cap$ interpersonal, 293 are AI $\cap$ ICT, and 9 are in all three. The category-level cosine similarities are ICT-AI 0.35, ICT-interpersonal 0.15, and interpersonal-AI 0.01, supporting the conceptual distinction between the categories. The AI category is then refined by a separate cosine-similarity threshold against the OECD AI taxonomy of Baruffaldi et al. (2020) (200 reference terms) using the all-MiniLM-L6-v2 sentence-transformer embedding of Reimers and Gurevych (2019): only AI-classified labels with embedding similarity $\geq \tau$ to the OECD reference are retained. The threshold $\tau = 0.50$ removes the generic terms (Python, troubleshoot, cybersecurity, alarm systems, electronics) that are common false positives and brings the AI mention rate from approximately 13% to approximately 3%.

Following classification, the three broad categories are mapped back to each job advertisement,

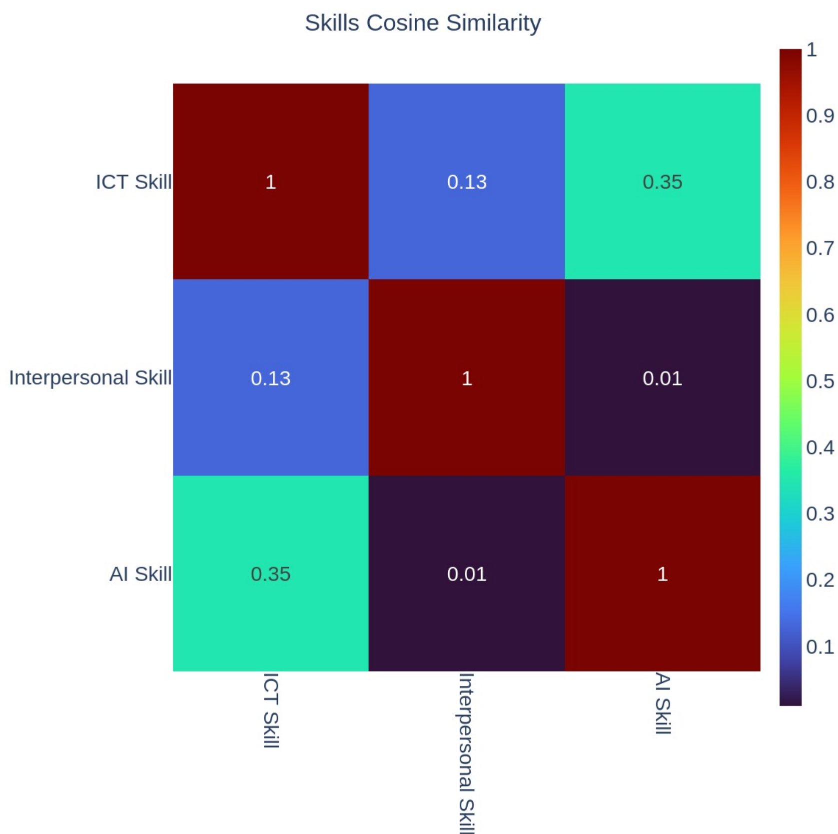

Figure A.1: Cosine similarity of the broader skill categories

yielding binary variables for whether a posting requires ICT, interpersonal, or AI skills, or a combination of them.

Classifier coverage. Of the 10,796 unique skills extracted, the GPT-4 classifier assigns approximately 3,000 to interpersonal, 2,500 to ICT, and 300 to AI; the remaining 4,000 residual skills span the numeracy, literacy, problem-solving, decision-making, creative, and manual categories of the broader ESCO taxonomy that are not the focus of the paper (Appendix A.4). The exclusion is uniform across time periods and occupations, so it does not bias the within-occupation, within-period premia.

Why these three skill categories. The substantive motivation is given in Section 1 (Introduction): ICT and interpersonal skills are the categories whose demand was most directly reshaped by the COVID-19 transition to remote and digitally-mediated work, and AI is central to the technological-change literature on the labour market.

Sample restrictions and analytic sample. Of the 83 million postings with skill information and a valid date in the April 2016 to December 2022 window, approximately 57.8 million report a wage (a 69.8% disclosure rate); when a single value is reported (rather than a min-max range), it is treated as both minimum and maximum. The sample is restricted to advertisements with wages above the 2016 annual minimum-wage floor (£13,936, for 40 hours per week and 52 weeks) and below the 99^{th} percentile of the wage distribution, to full-time positions, and to postings with non-missing SOC, ITL1, and the three primary skill variables. Adzuna (2025) could not aggregate the December 2019 data, and that month is dropped. These filters yield an analytic sample of approximately 41 million observations, of which approximately 22 million mention an ICT skill (the AI sub-sample used in the AI regressions of Section 4.2).

The UK as a high-disclosure setting. Around 70% of the 83M-row upstream sample carries explicit wage information, an unusually high disclosure rate in international perspective. Adrjan and Lydon (2024) document substantial cross-country heterogeneity in posted-wage frequency across eight advanced economies, and Batra et al. (2023) report that only around 14% of US online vacancies disclose any salary detail. The UK rate plausibly reflects a combination of stronger wage-transparency norms in recruitment practice, regulatory pressure from the Equality Act, and the Adzuna aggregation pipeline that re-publishes wage information when present in the source advertisement. The UK Adzuna corpus is therefore the closest currently-available analogue of a posted-wage administrative census for an advanced economy.

Selection into the wage-posted subsample. Whether a vacancy advertises a wage may depend on the skills demanded. A linear-probability regression of $\mathbf{1}\{\text{wage posted}\}$ on the three skill variables and year fixed effects, estimated on the raw upstream sample of 83 million advertisements (prior to any analytic-sample filter), gives $\hat{\beta}_{\text{AI}} = -0.059$ (AI vacancies are 5.9 percentage points less likely to post a wage; SE = 0.0002), $\hat{\beta}_{\text{Int}} = -0.060$, and $\hat{\beta}_{\text{ICT}} = -0.0023$ (all $p < 0.001$). The economic magnitudes are modest, around 70% of advertisements post a wage regardless of skill content, but the analytic sample slightly over-represents non-AI and non-interpersonal vacancies. The posted-wage premia in Section 4 should therefore be interpreted as conditional on the joint event of vacancy posting and wage disclosure, not as population-average skill returns.

A.3 Skill Extraction Example (Kanders and Sleeman (2021))

Title : Agile business analyst
Description = 'Support the Business to continually improve the business needs by working hand-in-hand with the Product Owner and Development teams. Work with product owners to write & develop clear, non-implementation specific epics, user stories and acceptance criteria. Interview product owners to understand as-is business processes and then develop customer-driven to-be processes. Progressively improve our demand side and analysis practices, focusing on waste elimination, demonstrating this improvement with hard data. Enforce and promote SCRUM disciplines. Should be able to organize and run daily stand ups. Promotes healthy team environment and removes impediments Ensures team is delivering / aligned on project vision and goals A strong background in retail financial services and a proven ability to quickly

understand the business strategy and objectives Significant experience of business analysis in an agile environment An IT background with solid grounding in technology is essential Experience of developing high quality user stories and acceptance criteria for multiple business processes having multiple product owners in an organization new to agile development methodology 3+ years in the role of Business Analyst Communication, group dynamics, collaboration and continuous improvement are - core being best practice driven Kanban practitioner, Scrum Certified or Six Sigma certification a plus 1 or higher in a related discipline from an academic institution; Masters a plus'

Table A.2: Detected Skills

Surface Form	Surface Form Type	Preferred Label	Entity	Predicted Q	Cluster 0	Cluster 1	Cluster 2
team work	manual	work as a team	10228	0.856	0.0	0.0	0.0
communication	label_pref	communication	1139	0.497	0.0	0.0	0.0
business analysis	label_pref	business analysis	5288	0.914	5.0	10.0	24.0
agile development	label_pref	Agile development	11555	0.919	5.0	10.0	24.0
business process	label_pref	business processes	8412	0.926	5.0	10.0	24.0
agile	label_alt	ICT project management methodologies	10023	0.797	5.0	10.0	24.0
scrum	label_alt	ICT project management methodologies	10023	0.874	5.0	10.0	24.0
continuous improvement	chunk_pref	create a work atmosphere of continuous improvement	8367	0.656	5.0	10.0	24.0
financial service	chunk_pref	offer financial services	11948	0.745	5.0	11.0	27.0
kanban	chunk_descr	continuous improvement philosophies	236	0.135	0.0	0.0	0.0

A.4 Example of classified skill

Table A.3: Skill Categories and Skills

Skill Category	Skills
ICT Skill Only	manage dispatch software systems, computer equipment, video-games functionalities, arrange audit, database
Interpersonal Skill Only	perform services in a flexible manner, customer service, sales activities, communication, German
ICT & Interpersonal Skill	stay up to date with social media, supply chain management, entertainment industry, offer financial services, internal auditing
ICT & AI Skill	alarm systems, control systems, cyber security, electronics, signal processing
ICT, Interpersonal & AI Skill	manage engineering project, climate change impact, electronic communication, prepare credit offers
Others	economics, attention to detail, work independently, develop energy policy

A.5 Word clouds for classified skills

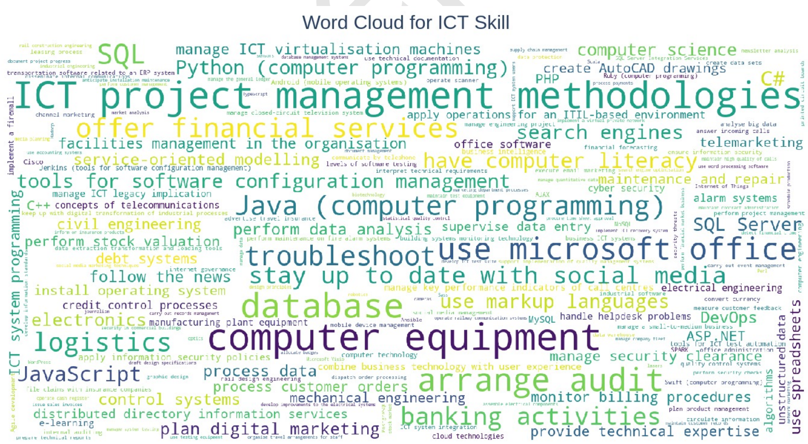

Figure A.2: Distribution of Extracted ICT Skills : This word cloud depicts the relative frequency of unique skills identified from job advertisements. The size of each term corresponds to its occurrence within the analysed job advertisement data.

A.6 Surface-form mechanics of the Nesta-ESCO pipeline

I use the methodology and algorithm provided by Kanders and Sleeman (2021) to extract skills from the job description of each advertisement. They adopt a top-down approach by using existing skill entities (labels, descriptions) as a reference, to extract skills from online job advertisements. It leverages the European Commission's Skill and Competences Organisation (ESCO)

Figure A.3: Distribution of Extracted Interpersonal Skills : This word cloud depicts the relative frequency of unique skills identified from job advertisements. The size of each term corresponds to its occurrence within the analysed job advertisement data.

framework as a comprehensive reference for skills and knowledge. ESCO categorises information for thousands of occupations and their associated skills, encompassing knowledge, abilities, and values. These are all referred to as "skill entities". Each skill entity has a "preferred label", "alternative labels", and a description. It can also link to broader or narrower skills and highlight occupations where the skill is important. To identify these skills in job postings, the model first analyses the text of labels and descriptions. NLP techniques are used to create "surface forms" - simpler terms that capture the essence of the skill entity. Two main types of surface forms are generated. The first involves cleaning preferred and alternative labels (removing punctuation, converting to lowercase, and simplifying word forms). The second extracts noun phrases from labels and descriptions using a library called spaCy. This is particularly useful for longer descriptions that provide context about the skill. To avoid ambiguity, the model prioritises surface forms derived from labels over descriptions, and those created using simpler cleaning methods.

Once generated, surface forms are evaluated for quality. This involves using metrics like Term Frequency - Inverse Document Frequency (tf-idf) to identify forms that might be more relevant to other skills. Additionally, a subset of forms is reviewed manually, and a supervised learning model is trained to predict the quality of the remaining ones. This model achieves high accuracy in filtering out low-quality surface forms, such as generic terms or job benefits.

Finally, filtered surface forms ($\sim$ 100,000) are compared with phrases from job descriptions for skill detection in job postings using the "Phrasematcher" feature in the spaCy library. Kanders and Sleeman (2021) further extend their work by grouping the surface forms into a three-level hierarchy of skills using clustering, by creating numerical representations of each form and then grouping similar forms together. An extended worked example is provided in Appendix A.3.

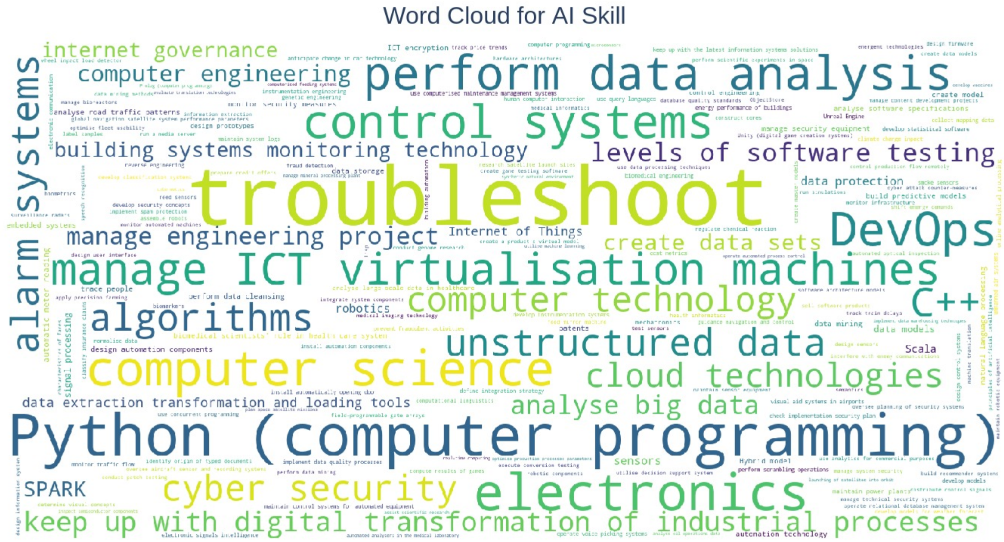

Figure A.4: Distribution of Extracted AI Skills : This word cloud depicts the relative frequency of unique skills identified from job advertisements. The size of each term corresponds to its occurrence within the analysed job advertisement data.

ESCO extraction: an illustration. Consider the most prevalent extracted skill, "communication". The Nesta-ESCO pipeline identifies the surface form via the SpaCy phrase matcher and maps it to the ESCO preferred label "communication". In contrast, the ESCO preferred label "perform services in a flexible manner" is matched whenever vacancy text mentions related variants ("perform services flexibly", "flexible service delivery"); the matcher picks up the partial surface form, looks up its parent in the ESCO taxonomy, and emits the preferred-label form. Two implications follow for Figure 2.2 in the main text. First, generic-sounding compound phrases surface prominently because the ESCO taxonomy encodes broad transversal competencies as full-phrase preferred labels. Second, a high-frequency umbrella concept such as "management" does not appear prominently in the figure because the ESCO taxonomy splits it across many more specific preferred labels ("business management", "project management", "operations management"), distributing the count across the taxonomy rather than concentrating it on the bare noun.

A.7 GPT-4 API call structure

The core aspect of the skill-classification approach involves the integration of the GPT-4 API. The skill-classification process proceeds as follows.

- Preparation Prompt: An initial prompt is constructed, that is *"Using the above definitions, classify the following."* This prompt sets the stage for the GPT-4 API to receive skill-specific inputs for classification.
- Skill Input: Individual skills to be classified are then presented as part of the overall prompt. This allows the GPT-4 API to analyze each skill in isolation based on the pre-

defined definitions. *Categorise skill '<skill>' as "ICT Skill", "Interpersonal Skill", "AI Skill". Print only 0 or 1 for each case in the format 'category: value'.*

- GPT-4 API Integration: The GPT-4 API uses its advanced language-processing capabilities to categorise the presented skill. It compares the skill to the pre-defined definitions of ICT, Interpersonal, and AI skills.
- Category Output: For each skill, the GPT-4 API delivers a binary classification in the format "category: value". Here, "category" represents one of the three predefined skill categories ("ICT Skill", "Interpersonal Skill", or "AI Skill"), and "value" is either 0 (not belonging to the category) or 1 (belonging to the category). This binary output provides a clear and concise classification for each input skill.

A.8 GPT-4 zero-shot classification: trade-offs

Zero-shot LLM classification trades manual coding cost against two sources of measurement error that bear on the economic interpretation of the estimated premia.

Classifier accuracy is below one. Validation on a random sample of 100 skills yields 92% (ICT), 96% (interpersonal), and 84% (AI) accuracy (Table A.1). Under the standard errors-in-variables benchmark for a binary regressor, the OLS coefficient is attenuated by roughly one minus the reliability ratio, so the AI coefficient is biased toward zero by roughly 16% and the ICT and interpersonal coefficients by less. The AI category is the most exposed because its content sits at the boundary between technical AI and broader software competency. The cosine-similarity refinement against the OECD AI taxonomy of Baruffaldi et al. (2020) addresses this margin: at $\tau = 0.50$ the retained vocabulary drops the generic terms, and the AI mention rate falls from about 13% to about 3%. Figure 4.2 reports the AI premium across the range of τ; Appendix B.3 re-estimates the premium with the Lightcast vocabulary in place of the OECD reference and confirms that the conclusion does not depend on the reference taxonomy.

The binary classification also discards information at the multi-membership margin. Some skills are intrinsically multifaceted, and a single-label output forces a hard partition that the ESCO taxonomy itself does not impose. The empirical impact is limited because the three target categories sit at a coarse level of aggregation: ICT versus interpersonal cross-overlap is modest (cosine similarity 0.15; Figure A.1), and AI is by construction a strict subset of ICT in the analytic sample. A multi-label or graded-membership scheme is a natural extension; under the present binary scheme the residual misclassification at the boundary is absorbed into the attenuation above.

Two further caveats apply to historical interpretation. The 2023-trained GPT-4 classifier may label skills as AI that were not recognised as such in 2016, producing a modest upward bias on the early-period AI mention rate; the cubic time control and the calendar-year fixed effects in the robustness checks absorb most of this. The pipeline is applied to the full posting rather than zoned to the requirements section, so mentions originating in firm description, employer branding, or benefits text contribute a weakly upward incidence bias; a transformer-based zoning classifier is left to future work.

A.9 Cross-source validation against LFS realised wages

The most informative external check on the Adzuna posted-wage data would be to estimate the same ICT, interpersonal, and AI premia on the UK Labour Force Survey and compare them to the present estimates. The LFS does not record the skill content of individual jobs, so a direct comparison of skill premia is not feasible. The wage-*level* variation can nonetheless be compared: the LFS reports realised hourly wages on a population-representative sample, which I aggregate to the same occupation-quarter cells as the Adzuna analytic sample. At the SOC-1 (major group) level, the Adzuna and LFS series correlate at $\rho = 0.945$ across 99 occupation-quarter cells (Figure A.5a); at the SOC-2 (sub-major group) level, $\rho = 0.910$ across 184 cells (Figure A.5b). The cross-occupation and cross-time variation in posted wages therefore tracks the corresponding variation in realised wages. The comparison speaks to the wage level rather than the skill return, but rules out the possibility that the within-occupation variation exploited in the regressions is purely an advertising-platform artefact.

SOC-1 major group ($\rho = +0.945$, 99 cells) SOC-2 sub-major group ($\rho = +0.910$, 184 cells)

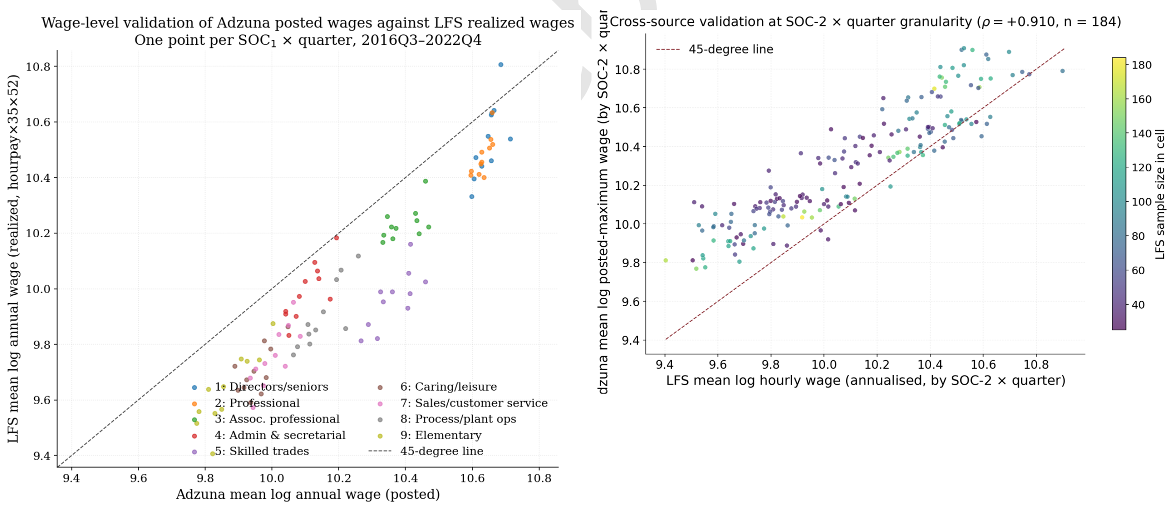

Figure A.5: LFS realised log hourly wages versus Adzuna posted log maximum wages, aggregated to occupation × calendar quarter (April 2016 to December 2022). Marker shading indicates the LFS cell sample size.

A.10 Adzuna corpus representativeness against ONS Textkernel adverts

Table A.4 compares the 83-million-row raw Adzuna corpus (April 2016 to December 2022) to the ONS Textkernel online job adverts series over the overlapping January 2017 to December 2022 window. The ONS reference shares are computed from Table 3 of the ONS *Labour demand* *by occupation* publication, summed across months and SOC-4 codes per ITL1 region (regions panel) and summed across months and regions per SOC-1 major group (occupations panel). For each Adzuna region and SOC-1 group the table also reports the prevalence of ICT, interpersonal and AI ($\tau = 0.50$) skill mentions, the share of postings with a non-missing posted wage, and the average posted maximum wage on the wage-disclosed subset. The headline numbers cited in Section 2 (*Representativeness against ONS* and *Wage disclosure across regions and occupations*) are derived directly from this table.

Table A.4: Adzuna corpus representativeness against ONS Textkernel online job adverts, by ITL1 region (top panel) and SOC-1 major group (bottom panel). All Adzuna columns are computed on the 83-million-row raw corpus; ONS reference shares cover the overlapping January 2017 to December 2022 window.

	% all posts		% AI	% ICT	% IP	% wage	Avg wage
	ONS	Adzuna	($\tau = 0.50$)			disclosed	(£k)
Regions (ITL1)							
East Midlands	6.3	5.9	2.09	51.0	88.9	74.1	32.1
East of England	8.9	8.9	2.84	51.3	89.4	71.3	34.1
London	22.7	20.8	3.80	59.6	90.8	69.5	46.6
North East	2.6	2.1	2.26	48.8	90.9	70.0	31.3
North West	10.0	9.0	2.33	52.5	90.9	73.0	33.3
Northern Ireland	1.5	0.8	3.67	62.6	91.6	48.5	33.2
Scotland	5.3	4.6	2.77	50.6	90.9	64.5	33.2
South East	15.6	16.4	2.79	52.9	90.3	72.9	35.6
South West	8.7	8.2	2.70	50.4	89.6	72.0	34.1
Wales	3.1	2.3	2.16	48.9	90.9	72.3	33.1
West Midlands	8.6	7.5	2.53	52.0	89.1	74.2	33.7
Yorkshire and The Humber	6.7	5.9	2.23	51.9	90.1	74.2	32.4
NULL	—	7.7	2.82	52.6	89.5	49.5	39.2
Occupations (SOC-1 major group)							
Managers, Directors and Senior Officials	6.9	9.2	2.33	56.0	95.6	67.0	48.0
Professional Occupations	28.6	33.2	4.91	58.3	90.5	70.0	46.8
Associate Professional Occupations	19.9	16.3	2.62	63.6	93.5	71.8	35.7
Administrative and Secretarial Occupations	10.5	7.6	1.07	67.4	92.7	74.1	25.4
Skilled Trades Occupations	8.3	7.6	2.93	44.2	76.2	74.0	33.0
Caring, Leisure and Other Service Occupations	7.8	4.8	0.47	20.0	95.8	73.5	20.6
Sales and Customer Service Occupations	4.5	5.4	0.38	48.7	99.1	65.7	22.9
Process, Plant and Machine Operatives	5.8	5.0	1.37	41.9	77.4	74.1	25.6
Elementary Occupations	7.7	5.1	0.45	31.6	83.2	66.3	18.9
NULL	—	5.8	2.37	48.0	89.1	56.0	38.3

Notes. Adzuna column reports the 82-million-row raw Adzuna corpus (April 2016 to December 2022). ONS column reports Textkernel online job adverts via the Office for National Statistics over the overlapping January 2017 to December 2022 window (ONS labour demand by occupation statistics, Table 3), summed across months and SOC-4 codes per region (regions panel) or summed across months and regions per SOC-1 major group (occupations panel). A posting is flagged as AI if classified as AI *and* its cosine similarity against the OECD AI taxonomy is ≥ 0.50. "% wage disclosed" is the share of postings in the group with a non-missing posted-wage value. Avg wage is the mean maximum posted wage in £k on the wage-disclosed subset. NULL rows report postings with missing region or SOC respectively.

A.11 OECD AI taxonomy (Baruffaldi et al. 2020)

The cosine-similarity refinement step described in Section 2 compares each extracted ESCO preferred label against the 200-term OECD AI taxonomy of Baruffaldi et al. (2020) using Sentence-BERT embeddings. The list below reports the 194 reference terms that appear as the nearest neighbour for at least one of the 10,796 extracted ESCO labels in the analytic sample; the remaining six taxonomy terms never appear as nearest neighbour in this corpus and are omitted for compactness. The full 200-term taxonomy is given in Baruffaldi et al. (2020) (Appendix B).

“3D Reconstruction”, “AIOps (Artificial Intelligence For IT Operations)”, “ANTLR”, “AWS SageMaker”, “Activity Recognition”, “AdaBoost (Adaptive Boosting)”, “Advanced Driver Assistance Systems”, “Advanced Robotics”, “Amazon Textract”, “Apache MADlib”, “Apache MXNet”, “Apache Mahout”, “Apache SINGA”, “Apache Spark”, “Applications Of Artificial Intelligence”, “Artificial General Intelligence”, “Artificial Intelligence”, “Artificial Intelligence Development”, “Artificial Intelligence Markup Language (AIML)”, “Artificial Intelligence Systems”, “Artificial Neural Networks”, “Association Rule Learning”, “Autoencoders”, “Automated Machine Learning”, “Autonomic Computing”, “Autonomous Cruise Control Systems”, “Autonomous System”, “Autonomous Vehicles”, “Azure Cognitive Services”, “Azure Machine Learning”, “Baidu”, “Boosting”, “CHi-Squared Automatic Interaction Detection (CHAID)”, “CUDNN”, “Caffe”, “Caffe2”, “Chainer (Deep Learning Framework)”, “Chatbot”, “Classification And Regression Tree (CART)”, “Cluster Analysis”, “Cognitive Automation”, “Cognitive Computing”, “Cognitive Robotics”, “Collaborative Filtering”, “Computational Intelligence”, “Computational Linguistics”, “Computer Vision”, “Confusion Matrix”, “Contextual Image Classification”, “Convolutional Neural Networks”, “Cortana”, “Cyber-Physical Systems”, “Dask (Software)”, “Data Classification”, “Dbscan”, “Decision Models”, “Decision Tree Learning”, “DeepSpeech”, “Deeplearning4j”, “Dialog Systems”, “Digital Image Processing”, “Dimensionality Reduction”, “Dlib (C++ Library)”, “Ensemble Methods”, “Evolutionary Programming”, “Expectation Maximization Algorithm”, “Expert Systems”, “Eye Tracking”, “Face Detection”, “Facial Recognition”, “Feature Engineering”, “Feature Extraction”, “Feature Learning”, “Feature Selection”, “Fuzzy Logic”, “Gaussian Process”, “Genetic Algorithm”, “Google AutoML”, “Google Cloud ML Engine”, “Gradient Boosting”, “Guidance Navigation And Control Systems”, “H2O.ai”, “Handwriting Recognition”, “Hidden Markov Model”, “Hugging Face (NLP Framework)”, “Hugging Face Transformers”, “Hyperparameter Optimization”, “IPSoft Amelia”, “Image Analysis”, “Image Matching”, “Image Processing”, “Image Recognition”, “Image Segmentation”, “Image Sensor”, “ImageNet”, “Inference Engine”, “Intelligent Agent”, “Intelligent Control”, “Intelligent Software Assistant”, “Intelligent Systems”, “Intelligent Virtual Assistant”, “Interactive Kiosk”, “Kaldi”, “Keras (Neural Network Library)”, “Kernel Methods”, “Knowledge-Based Configuration”, “Knowledge-Based Systems”, “Kubeflow”, “LIBSVM”, “Latent Dirichlet Allocation”, “Lexalytics”, “Light Detection And Ranging (LiDAR)”, “Long Short-Term Memory (LSTM)”, “MLOps (Machine Learning Operations)”, “MLflow”, “Machine Learning Algorithms”, “Machine Translation”, “Machine Vision”, “Markov Chain”, “Meta Learning”, “Microsoft Cognitive Toolkit (CNTK)”, “Microsoft LUIS”, “Motion Analysis”, “Motion Planning”, “Multi-Agent Systems”, “Naive Bayes”, “Natural Language Generation”, “Natural Language Processing Systems”, “Natural Language Programming”, “Natural Language Toolkits”, “Natural Language Understanding”, “Natural Language User Interface”, “Nearest Neighbour Algorithm”, “Nvidia Jetson”, “Object Recognition”, “OmniPage”, “OpenCV”, "OpenNLP", "OpenVINO", "Optical Character Recognition (OCR)", "PaddlePaddle", "Path Analysis", "Path Finding", "Perceptron", "Pose Estimation", "PredictionIO", "PyBrain", "Random Forest Algorithm", "RealSense", "Recommendation Engine", "Recommender Systems", "Recurrent Neural Network (RNN)", "Reinforcement Learning", "Remote Sensing", "Robot Framework", "Robot Operating Systems", "Robotic Automation Software", "Robotic Liquid Handling Systems", "Robotic Programming", "Robotic Systems", "SLAM Algorithms (Simultaneous Localization And Mapping)", "Scikit-learn (Machine Learning Library)", "Screen Reader", "Semantic Analysis", "Semantic Interpretation For Speech Recognition", "Semantic Parsing", "Semantic Search", "Semi-Supervised Learning", "Sentiment Analysis", "Seq2Seq", "Servomotor", "Soft Computing", "Sorting Algorithm", "Speech Recognition", "Speech Recognition Software", "Statistical Language Acquisition", "Supervised Learning", "Support Vector Machine", "TensorFlow", "Test Datasets", "Text Mining", "Tokenization", "Torch (Machine Learning)", "Training Datasets", "Transfer Learning", "Unmanned Aerial Systems (UAS)", "Unsupervised Learning", "Voice Interaction", "Voice User Interface", "Vowpal Wabbit", "Word Embedding", "XGBoost", "fastText", "mlpack (C++ Library)".

A.12 Lightcast AI/ML Open Skill Taxonomy (Bone et al. 2025)

For the vocabulary-sensitivity robustness exercise in Section 4.2, the AI flag is reconstructed using the Lightcast Open Skills AI/ML subcategory in place of the OECD reference. The 157 Lightcast skills below are reproduced verbatim from Bone et al. (2025) (Appendix 1, Section 1.1), themselves drawn from the Lightcast Open Skills taxonomy (`https://lightcast.io/open-skills/categories`).

"Artificial Intelligence", "Reinforcement Learning", "Voice User Interface", "Machine Learning Model Training", "Deep Learning Methods", "Computational Intelligence", "Dialog Systems", "Transformer (Machine Learning Model)", "Intelligent Systems", "Scikit-Learn (Python Package)", "Knowledge-Based Configuration", "AIOps (Artificial Intelligence For IT Operations)", "Language Model", "AdaBoost (Adaptive Boosting)", "OpenAI Gym", "Dlib (C++ Library)", "Google Cloud ML Engine", "PyTorch Lightning", "mlpack (C++ Library)", "Generative Adversarial Networks", "Recommender Systems", "MLOps (Machine Learning Operations)", "Knowledge Engineering", "OpenCV", "Theano (Software)", "Open Neural Network Exchange (ONNX)", "Intelligent Control", "Text-To-Speech", "Attention Mechanisms", "Game Ai", "H2O.ai", "AI Copywriting", "Adversarial Machine Learning", "OpenVINO", "Pydata", "Seq2Seq", "Google Bard", "IPSoft Amelia", "Apache SINGA", "Caffe (Framework)", "Chatbot", "Apache Mahout", "Dask (Software)", "Keras (Neural Network Library)", "Autoencoders", "Long Short-Term Memory (LSTM)", "Azure Cognitive Services", "AI/ML Inference", "Applications Of Artificial Intelligence", "Cognitive Computing", "Amazon Alexa", "Watson Studio", "Explainable AI (XAI)", "Programmatic Media Buying", "Generative Artificial Intelligence", "Cognitive Robotics", "Bot Framework", "Kubeflow", "Fast.ai", "AWS Certified Machine Learning Specialty", "Semi-Supervised Learning", "PaddlePaddle", "Meta Learning", "Google AutoML", "Feature Learning", "Weka", "Transfer Learning", "Swarm Intelligence", “Watson Conversation”, “Prompt Engineering”, “Stable Diffusion”, “TensorFlow”, “Boosting”, “AWS SageMaker”, “Machine Learning Methods”, “Feature Extraction”, “Caffe2”, “Feature Selection”, “Training Datasets”, “Artificial Intelligence Markup Language (AIML)”, “Intelligent Agent”, “Gesture Recognition”, “PyTorch (Machine Learning Library)”, “Machine Learning Algorithms”, “MLflow”, “Feature Engineering”, “Unsupervised Learning”, “Ethical AI”, “Artificial Intelligence Risk”, “Supervised Learning”, “Cudnn”, “Random Forest Algorithm”, “Genetic Algorithm”, “Speech Synthesis”, “Voice Interaction”, “Oracle Autonomous Database”, “Nvidia Jetson”, “Activity Recognition”, “Perceptron”, “Deeplearning4j”, “Operationalizing AI”, “Multi-Agent Systems”, “Voice Assistant Technology”, “Cognitive Automation”, “Knowledge-Based Systems”, “Speech Recognition Software”, “Hidden Markov Model”, “Microsoft Cognitive Toolkit (CNTK)”, “Kaldi”, “Artificial Intelligence Development”, “Expert Systems”, “Inference Engine”, “Intelligent Virtual Assistant”, “Objective Function”, “DALL-E Image Generator”, “Deep Learning”, “General-Purpose Computing On Graphics Processing Units”, “Azure Machine Learning”, “Variational Autoencoders”, “Association Rule Learning”, “Machine Learning Model Monitoring And Evaluation”, “Test Datasets”, “ChatGPT”, “K-Nearest Neighbors Algorithm”, “3D Reconstruction”, “Loss Functions”, “LightGBM”, “Cortana”, “OmniPage”, “Machine Learning”, “Intelligent Automation”, “Recurrent Neural Network (RNN)”, “Artificial Neural Networks”, “Convolutional Neural Networks”, “Torch (Machine Learning)”, “Baidu”, “Support Vector Machine”, “Amazon Textract”, “Gradient Boosting”, “Collaborative Filtering”, “Embedded Intelligence”, “Automated Machine Learning”, “Apache MXNet”, “ModelOps”, “Ensemble Methods”, “Kernel Methods”, “Deck.gl”, “Microsoft LUIS”, “Confusion Matrix”, “Natural Language User Interface”, “Xgboost”, “Artificial Intelligence Systems”, “Reasoning Systems”, “Large Language Modeling”, “Sorting Algorithm”, “Interactive Kiosk”, “Soft Computing”.

B Robustness Checks

B.1 Wage Premium of the Skill Bundle: Robustness

I re-estimate the joint ICT and interpersonal bundle of Section 4.1.3 across an FE grid that varies occupation from SOC-1 to SOC-3 and location between ITL-1 and TTWA, for both the maximum and minimum posted wage. The cell estimates attenuate monotonically as occupation FE becomes finer and are essentially invariant to the choice between ITL-1 and TTWA. Tables B.1 and B.2 report the estimates; Figures B.1 and B.2 show the same coefficients graphically.

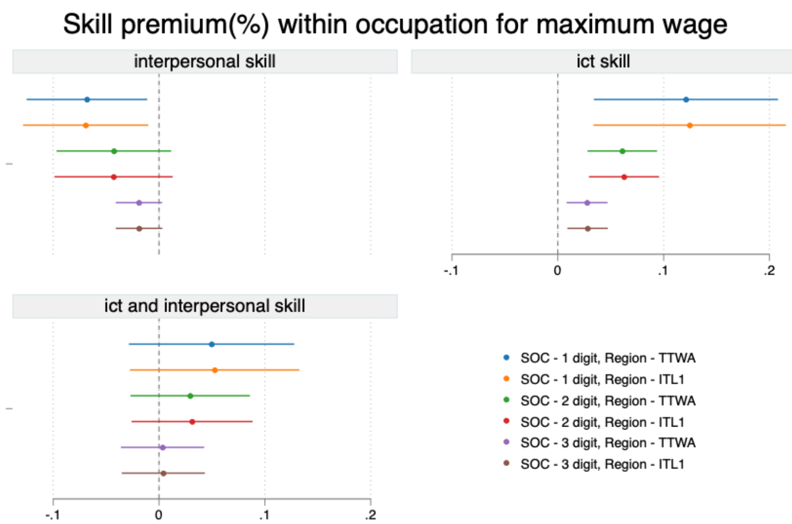

Figure B.1: Skill bundle wage premium for the maximum offered wage, estimated across the SOC-1 to SOC-3 occupation grid paired with ITL-1 or TTWA location absorption. Markers report the three saturated cell coefficients (ICT only; interpersonal only; ICT × interpersonal) with 95% confidence intervals.

Table B.1: Skill bundle wage premium for maximum wage offered

	log of maximum wage offered					
ict_skill=0 × interpersonal_skill=1	-0.0681**	-0.0693**	-0.0427	-0.0429	-0.0190*	-0.0189*
	(0.0247)	(0.0256)	(0.0235)	(0.0242)	(0.00949)	(0.00954)
ict_skill=1 × interpersonal_skill=0	0.121**	0.125**	0.0609***	0.0626***	0.0277**	0.0282***
	(0.0377)	(0.0394)	(0.0142)	(0.0144)	(0.00840)	(0.00828)
ict_skill=1 × interpersonal_skill=1	0.0497	0.0525	0.0294	0.0312	0.00333	0.00403
	(0.0339)	(0.0347)	(0.0245)	(0.0248)	(0.0171)	(0.0170)
Constant	10.33***	10.33***	10.33***	10.33***	10.33***	10.33***
	(0.0184)	(0.0186)	(0.0172)	(0.0173)	(0.0188)	(0.0186)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes
SOC FE - 1 Digit	Yes	Yes	No	No	No	No
SOC FE - 2 Digit	No	No	Yes	Yes	No	No
SOC FE - 3 Digit	No	No	No	No	Yes	Yes
Location FE - ITL1	No	Yes	No	Yes	No	Yes
Location FE - TTWA	Yes	No	Yes	No	Yes	No
Observations	41095009	41095009	41095009	41095009	37500231	37500231
r2	0.352	0.347	0.387	0.383	0.461	0.459
Standard Errors	clustered	clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: Skill bundle wage premium for minimum wage offered

	log of minimum wage offered					
ict_skill=0 × interpersonal_skill=1	-0.0732**	-0.0743**	-0.0481*	-0.0483*	-0.0266**	-0.0265**
	(0.0244)	(0.0253)	(0.0239)	(0.0245)	(0.0103)	(0.0103)
ict_skill=1 × interpersonal_skill=0	0.107**	0.110**	0.0447***	0.0462***	0.0128	0.0132*
	(0.0369)	(0.0384)	(0.0117)	(0.0118)	(0.00702)	(0.00684)
ict_skill=1 × interpersonal_skill=1	0.0476	0.0503	0.0228	0.0245	-0.00315	-0.00247
	(0.0341)	(0.0349)	(0.0230)	(0.0232)	(0.0166)	(0.0165)
Constant	10.20***	10.20***	10.20***	10.20***	10.20***	10.20***
	(0.0177)	(0.0180)	(0.0166)	(0.0167)	(0.0184)	(0.0183)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes
SOC FE - 1 Digit	Yes	Yes	No	No	No	No
SOC FE - 2 Digit	No	No	Yes	Yes	No	No
SOC FE - 3 Digit	No	No	No	No	Yes	Yes
Location FE - ITL1	No	Yes	No	Yes	No	Yes
Location FE - TTWA	Yes	No	Yes	No	Yes	No
Observations	41095009	41095009	41095009	41095009	37500231	37500231
r2	0.341	0.337	0.382	0.378	0.455	0.453
Standard Errors	clustered	clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

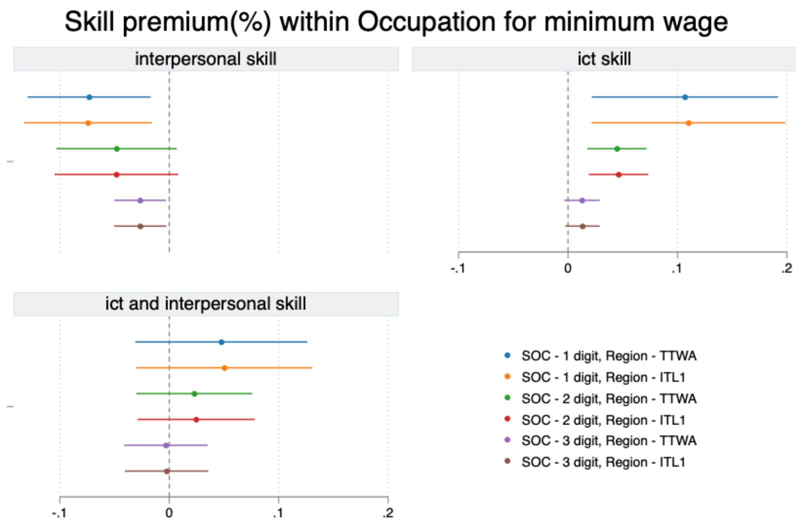

Figure B.2: Skill bundle wage premium for the minimum offered wage, estimated across the SOC-1 to SOC-3 occupation grid paired with ITL-1 or TTWA location absorption. Markers report the three saturated cell coefficients (ICT only; interpersonal only; ICT × interpersonal) with 95% confidence intervals.

B.2 Wage Premium of AI Skill: Robustness

I re-estimate the AI premium of Section 4.2 across an FE grid that varies occupation from SOC-1 to SOC-4 and location between ITL-1 and TTWA, for both the maximum and minimum wage. The premium is positive and significant in every specification, and the magnitude declines monotonically with finer SOC granularity, indicating that part of the unconditional AI-wage association reflects between-detailed-occupation sorting. The minimum-wage estimates track the maximum-wage estimates closely. The sample size in Tables B.3 and B.4 (around 19.3 million) corresponds to the legacy AI variable before the cosine-similarity refinement at $\tau = 0.50$; the headline specification in Section 4.2 uses the refined AI skill on the same ICT-conditional sample.

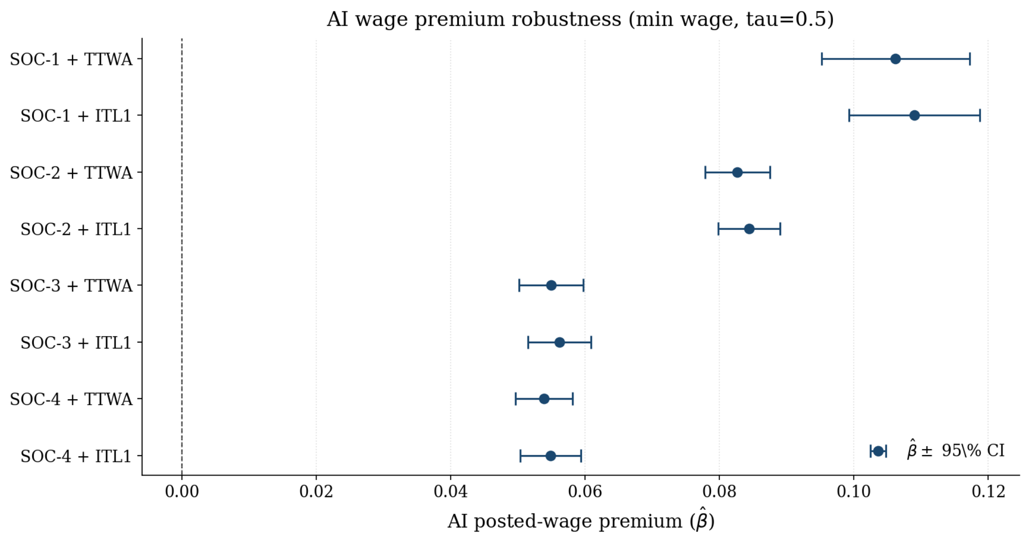

Figure B.3: AI posted-wage premium $\hat{\beta}_{\mathrm{AI}}$ at $\tau = 0.50$ on the minimum offered wage, across the SOC-1 to SOC-4 occupation grid paired with TTWA or ITL-1 location absorption. 95% confidence intervals.

Table B.3: AI wage premium for maximum wage offered

	log of maximum wage offered							
ai_skill	0.118***	0.121***	0.0825***	0.0843***	0.0660***	0.0673***	0.0649***	0.0662***
	(0.0222)	(0.0237)	(0.00744)	(0.00761)	(0.00676)	(0.00683)	(0.00658)	(0.00673)
Constant	10.39***	10.39***	10.39***	10.39***	10.39***	10.39***	10.39***	10.39***
	(0.0142)	(0.0150)	(0.0124)	(0.0129)	(0.0142)	(0.0146)	(0.0132)	(0.0136)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SOC FE - 1 Digit	Yes	Yes	No	No	No	No	No	No
SOC FE - 2 Digit	No	No	Yes	Yes	No	No	No	No
SOC FE - 3 Digit	No	No	No	No	Yes	Yes	No	No
SOC FE - 4 Digit	No	No	No	No	No	No	Yes	Yes
Location FE - ITL1	No	Yes	No	Yes	No	Yes	No	Yes
Location FE - TTWA	Yes	No	Yes	No	Yes	No	Yes	No
Observations	19328560	19328560	19328560	19328560	19328560	19328560	19328560	19328560
r2	0.380	0.373	0.404	0.399	0.451	0.447	0.467	0.463
Standard Errors	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.4: AI wage premium for minimum wage offered

	log of minimum wage offered							
ai_skill	0.111***	0.114***	0.0812	0.0828***	0.0638***	0.0649***	0.0618***	0.0629***
	(0.0172)	(0.0184)	(.)	(0.00680)	(0.00661)	(0.00671)	(0.00640)	(0.00654)
Constant	10.25***	10.25***	10.25	10.25***	10.25***	10.25***	10.25***	10.25***
	(0.0114)	(0.0122)	(.)	(0.0106)	(0.0127)	(0.0132)	(0.0115)	(0.0120)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SOC FE - 1 Digit	Yes	Yes	No	No	No	No	No	No
SOC FE - 2 Digit	No	No	Yes	Yes	No	No	No	No
SOC FE - 3 Digit	No	No	No	No	Yes	Yes	No	No
SOC FE - 4 Digit	No	No	No	No	No	No	Yes	Yes
Location FE - ITL1	No	Yes	No	Yes	No	Yes	No	Yes
Location FE - TTWA	Yes	No	Yes	No	Yes	No	Yes	No
Observations	19328560	19328560	19328560	19328560	19328560	19328560	19328560	19328560
r2	0.364	0.359	0.389	0.384	0.433	0.430	0.448	0.444
Standard Errors	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B.3 AI vocabulary sensitivity: Lightcast versus OECD reference

The headline AI skill is defined by retaining the GPT-4 AI-classified ESCO labels whose Sentence-BERT embedding has cosine similarity $\geq \tau = 0.50$ against the 200-term OECD AI taxonomy of Baruffaldi et al. (2020) (Appendix A.11). To see how sensitive the AI premium is to the choice of reference vocabulary, I re-estimate the premium across the same range of the cosine-similarity threshold τ with Bone et al.'s 157-term Lightcast AI/ML Open Skill Taxonomy (Appendix A.12; Bone et al. 2025, Appendix 1) replacing the OECD reference. For each posting, the Lightcast cosine score is the cosine similarity between the posting's nearest OECD AI term and its nearest neighbour in the Lightcast vocabulary, computed with the same all-MiniLM-L6-v2 model. The re-estimation then mirrors the OECD specification at the row-level threshold τ, with the same SOC-2 occupation and TTWA location fixed effects and three-way clustered standard errors, on the ICT-conditional analytic sample.

Figure B.4 overlays the two threshold curves. At low τ (broad vocabulary) the two curves coincide at approximately 8.6%. At the headline $\tau = 0.50$, the Lightcast vocabulary returns an AI premium of approximately 7.2% versus the OECD reference's 9.1%, a vocabulary-driven gap of about 1.9 percentage points. The Lightcast curve is flatter at higher τ, because the Lightcast vocabulary includes generic mentions ("Artificial Intelligence", "Machine Learning") that the OECD-refined definition removes. Across $\tau \in [0.30, 0.80]$, the two curves remain within a narrow 5% to 11% band, so the conclusion that AI carries a meaningful posted-wage premium is invariant to the choice of reference vocabulary.

B.4 External-validity replication: matched-specification AI premium

The headline AI premium of 9.1% in Section 4.2 is estimated on the ICT-conditional subsample with SOC-2 occupation, TTWA location, and a cubic time control. Bone et al. (2025) estimate UK posted-wage premia on a different vacancy source (Lightcast), with a different specification (SOC-1, SIC-1, NUTS-1, and year fixed effects), and on the full UK Lightcast sample without an ICT-conditional restriction, reporting an AI coefficient of $\hat{\beta} \approx 0.23$ to 0.26. To see whether the gap between the headline 9.1% and Bone et al.'s 23% to 30% reflects data-source or specification differences, I re-estimate the AI premium on the Adzuna analytic sample under the closest replication of their specification that the Adzuna fields permit: the full analytic sample (no ICT-conditional restriction), SOC-1, ITL-1, and year fixed effects. Education and experience information is sparsely populated at the posting level in Adzuna, so including those controls would substantially shrink the analytic sample toward the Bone-equivalent subsample on which education- and experience-controlled estimates can be obtained; the matched specification here preserves the larger sample at the cost of the controls. Standard errors are three-way clustered on (year-month, SOC-1, ITL-1). Estimates are reported for both the OECD 200-term AI taxonomy and the Lightcast 157-term AI/ML vocabulary at $\tau = 0.50$.

Under the Bone-matched specification, the OECD AI premium is 21.4% and the Lightcast AI

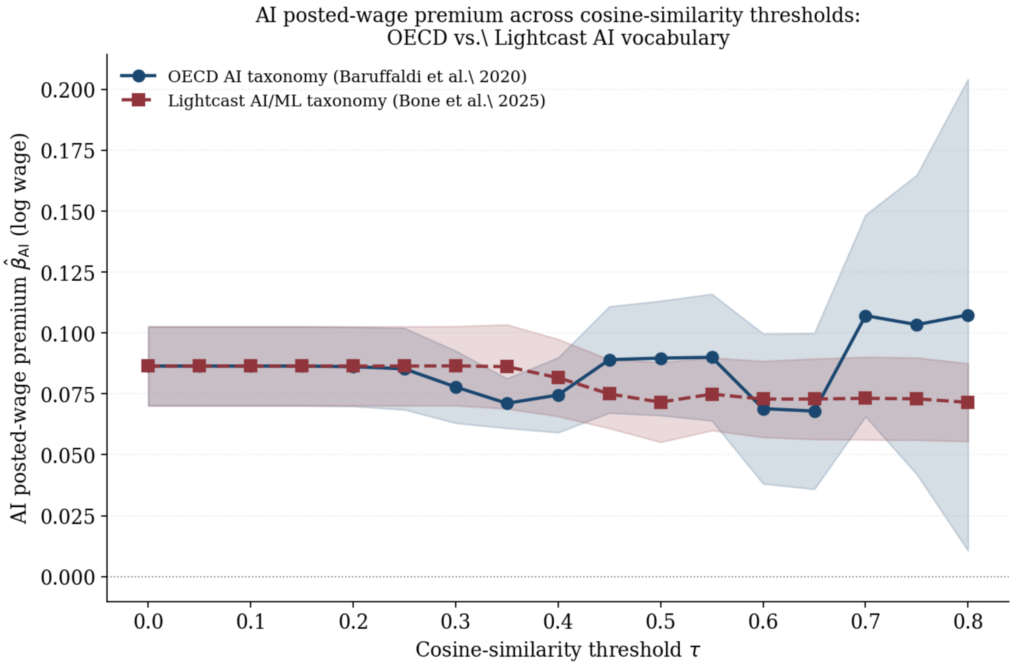

Figure B.4: AI posted-wage premium $\hat{\beta}_{\mathrm{AI}}$ with 95% confidence intervals across the cosine-similarity threshold τ, under two reference vocabularies: OECD AI taxonomy (Baruffaldi et al. 2020, 200 terms; solid navy line) and Lightcast AI/ML Open Skill Taxonomy (Bone et al. 2025, 157 terms; dashed maroon line). Each point is the AI coefficient in the headline log-wage regression with SOC-2 and TTWA fixed effects absorbed, 3-way CGM clustered standard errors, estimated on a 25% random subsample of the 22M ICT-conditional analytic sample. Shaded bands are 95% CIs.

Table B.5: AI posted-wage premium under Bone et al. (2025)-matched specification on the Adzuna analytic sample. The headline (ICT-conditional, SOC-2 + TTWA + cubic time) row is reported for comparison.

Specification	Vocabulary	$\hat{\beta}$ (SE)	Premium
Adzuna: SOC-2 + TTWA + cubic time, ICT	OECD	0.091 (0.010)	9.1%
Bone-matched: SOC-1 + ITL-1 + year FE, full	OECD	0.194 (0.029)	21.4%
Bone-matched: SOC-1 + ITL-1 + year FE, full	Lightcast	0.170 (0.044)	18.5%
Bone et al. (2025): UK Lightcast	Lightcast	—	+23% to +30%

3-way Cameron-Gelbach-Miller clustered SE on (year-month, SOC-1, ITL-1) in parentheses; all coefficients significant at $p < 0.001$.

premium is 18.5%, both within the 23% to 30% range that Bone et al. (2025) report on UK Lightcast data. The headline 9.1% is therefore not an artefact of the Adzuna corpus: matching the specification to Bone et al.'s brings the two corpora into broad agreement. The gap between the headline and Bone et al.'s estimate is driven by fixed-effect granularity (SOC-2 and TTWA versus SOC-1 and ITL-1) and the ICT-conditional sample restriction, not by the data source. The vocabulary choice (OECD versus Lightcast) accounts for roughly three percentage points of within-spec variation, in line with the OECD-versus-Lightcast overlay across the range of τ

in Appendix B.3.

B.5 Impact of COVID-19 on Skill Premia: Robustness

I re-estimate the COVID-interaction specifications of Section 4.3 across an FE grid that varies occupation from SOC-1 to SOC-3 and location between ITL-1 and TTWA, for both the maximum and minimum wage and separately for each skill. The level premia decline monotonically with finer SOC granularity and are essentially invariant to the choice between ITL-1 and TTWA. The COVID and post-COVID interactions remain small throughout, with only the ICT $\times$ COVID interaction approaching economic significance at coarser SOC absorptions. The minimum-wage estimates track the maximum-wage estimates throughout.

B.5.1 ICT Skill

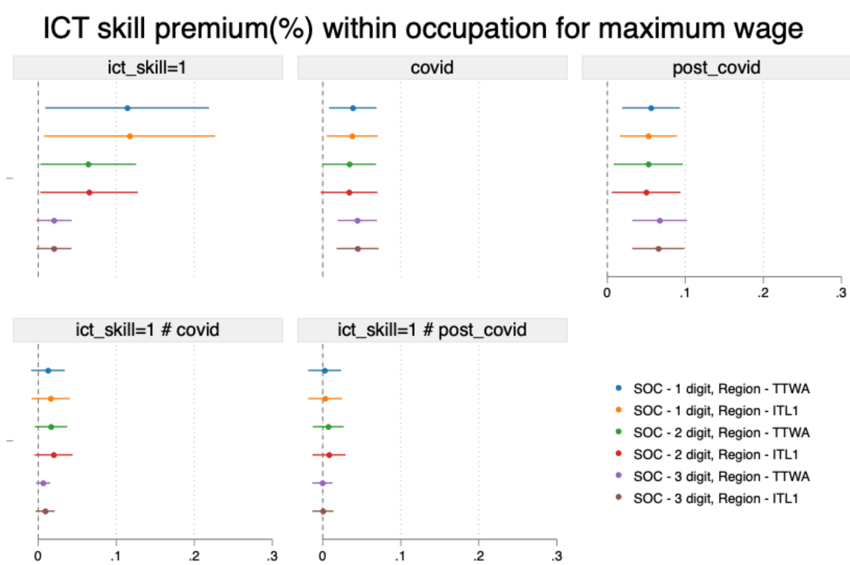

Figure B.5: Wage premium of ICT skill during COVID-19 for maximum wage offered

Table B.6: ICT skill wage premium and COVID-19 for maximum wage offered

	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
ict_skill=1	0.114**	0.117**	0.0641**	0.0652**	0.0202*	0.0201*
	(0.0455)	(0.0475)	(0.0266)	(0.0271)	(0.00986)	(0.00980)
covid	0.0385**	0.0376**	0.0341*	0.0338*	0.0442***	0.0447***
	(0.0132)	(0.0143)	(0.0148)	(0.0158)	(0.0110)	(0.0117)
post_covid	0.0560***	0.0527**	0.0526**	0.0498**	0.0671***	0.0654***
	(0.0161)	(0.0158)	(0.0192)	(0.0191)	(0.0152)	(0.0145)
ict_skill=1 × covid	0.0124	0.0159	0.0163	0.0198*	0.00620	0.00891
	(0.00935)	(0.0107)	(0.00903)	(0.0105)	(0.00398)	(0.00528)
ict_skill=1 × post_covid	0.00235	0.00305	0.00706	0.00806	-0.000425	0.000233
	(0.00915)	(0.00941)	(0.00864)	(0.00918)	(0.00561)	(0.00587)
Constant	10.27***	10.27***	10.29***	10.29***	10.31***	10.31***
	(0.0268)	(0.0282)	(0.0170)	(0.0175)	(0.0129)	(0.0133)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes
SOC FE - 1 Digit	Yes	Yes	No	No	No	No
SOC FE - 2 Digit	No	No	Yes	Yes	No	No
SOC FE - 3 Digit	No	No	No	No	Yes	Yes
Location FE - ITL1	No	Yes	No	Yes	No	Yes
Location FE - TTWA	Yes	No	Yes	No	Yes	No
Observations	41095009	41095009	41095009	41095009	37500231	37500231
r2	0.350	0.345	0.387	0.383	0.461	0.459
Standard Errors	clustered	clustered	clustered	clustered	clustered	clustered

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.7: ICT skill wage premium and COVID-19 for minimum wage offered

	log_wages_min	log_wages_min	log_wages_min	log_wages_min	log_wages_min	log_wages_min
ict_skill=1	0.116**	0.119**	0.0624**	0.0635**	0.0198**	0.0198**
	(0.0458)	(0.0476)	(0.0244)	(0.0248)	(0.00782)	(0.00771)
covid	-0.00132	-0.00227	-0.00444	-0.00490	0.00271	0.00290
	(0.00879)	(0.00970)	(0.0101)	(0.0107)	(0.00415)	(0.00486)
post_covid	0.0335**	0.0306*	0.0327*	0.0303*	0.0444***	0.0430***
	(0.0135)	(0.0136)	(0.0155)	(0.0155)	(0.0106)	(0.00993)
ict_skill=1 × covid	0.00475	0.00757	0.00893	0.0118	0.000947	0.00309
	(0.00860)	(0.00971)	(0.00928)	(0.0106)	(0.00568)	(0.00661)
ict_skill=1 × post_covid	0.00778	0.00836	0.0119	0.0128	0.00633	0.00688
	(0.0135)	(0.0138)	(0.0136)	(0.0142)	(0.0119)	(0.0122)
Constant	10.14***	10.14***	10.16***	10.16***	10.18***	10.18***
	(0.0272)	(0.0286)	(0.0161)	(0.0166)	(0.0120)	(0.0124)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes
SOC FE - 1 Digit	Yes	Yes	No	No	No	No
SOC FE - 2 Digit	No	No	Yes	Yes	No	No
SOC FE - 3 Digit	No	No	No	No	Yes	Yes
Location FE - ITL1	No	Yes	No	Yes	No	Yes
Location FE - TTWA	Yes	No	Yes	No	Yes	No
Observations	41095009	41095009	41095009	41095009	37500231	37500231
r2	0.339	0.335	0.381	0.378	0.455	0.453
Standard Errors	clustered	clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

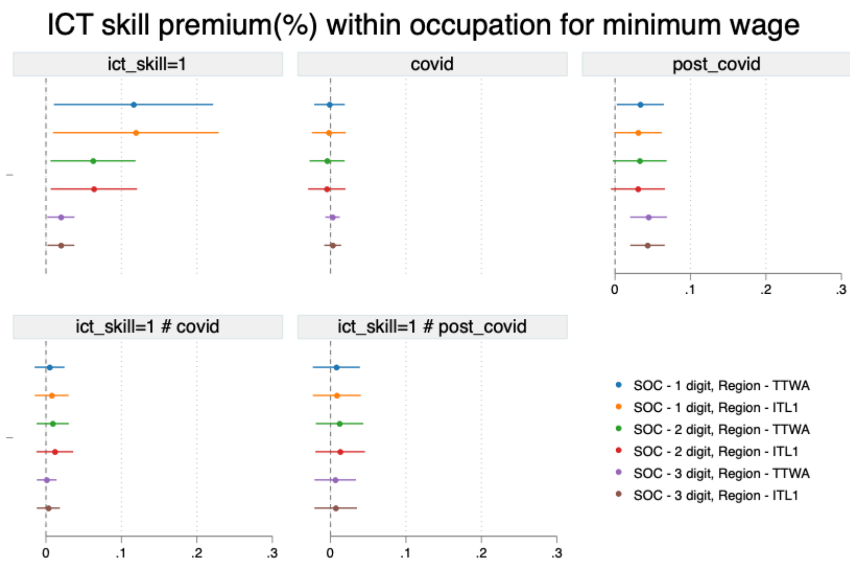

Figure B.6: Wage premium of ICT skill during COVID-19 for minimum wage offered

B.5.2 Interpersonal Skill

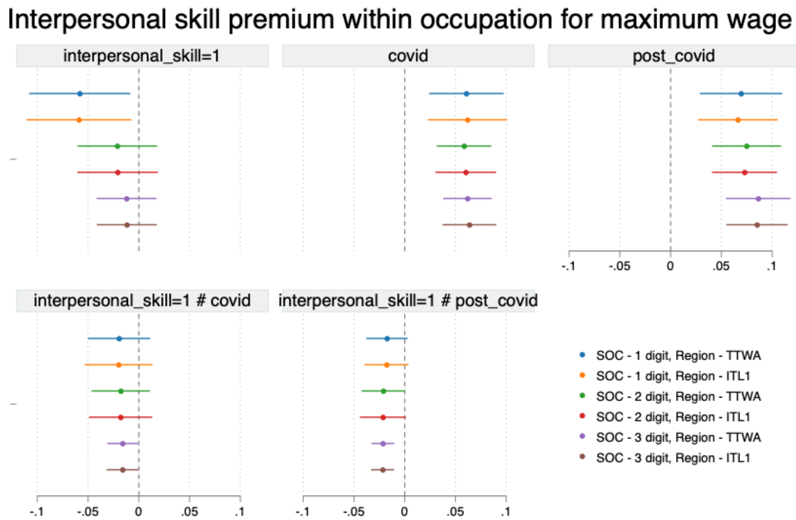

Figure B.7: Wage premium of interpersonal skill during COVID-19 for maximum wage offered

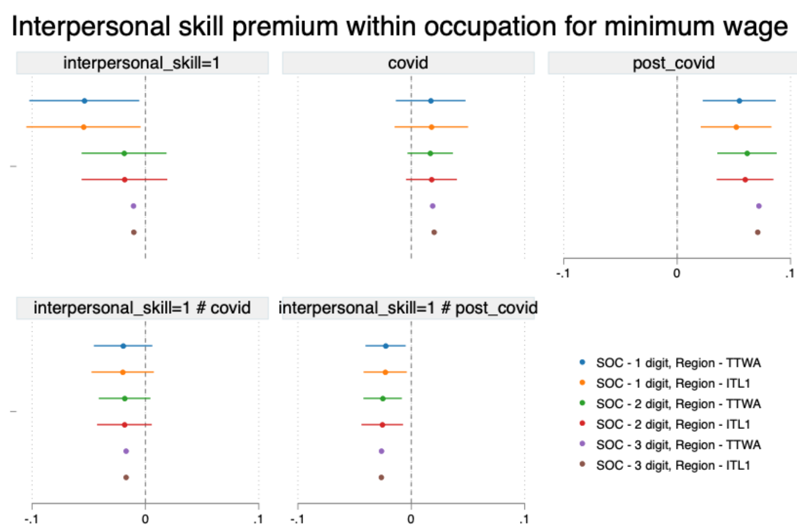

Figure B.8: Wage premium of interpersonal skill during COVID-19 for minimum wage offered

Table B.8: Interpersonal skill wage premium and COVID-19 for maximum wage offered

	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
interpersonal_skill=1	-0.0583**	-0.0591**	-0.0213	-0.0209	-0.0121	-0.0118
	(0.0216)	(0.0224)	(0.0170)	(0.0172)	(0.0128)	(0.0128)
covid	0.0607***	0.0618***	0.0585***	0.0601***	0.0618***	0.0637***
	(0.0159)	(0.0169)	(0.0116)	(0.0129)	(0.0104)	(0.0115)
post_covid	0.0694***	0.0662***	0.0749***	0.0728***	0.0864***	0.0851***
	(0.0176)	(0.0170)	(0.0148)	(0.0139)	(0.0138)	(0.0131)
interpersonal_skill=1 × covid	-0.0195	-0.0199	-0.0178	-0.0180	-0.0160**	-0.0161**
	(0.0133)	(0.0145)	(0.0125)	(0.0135)	(0.00657)	(0.00685)
interpersonal_skill=1 × post_covid	-0.0175*	-0.0179*	-0.0211*	-0.0215*	-0.0215***	-0.0217***
	(0.00884)	(0.00938)	(0.00931)	(0.00985)	(0.00484)	(0.00493)
Constant	10.38***	10.38***	10.34***	10.34***	10.33***	10.33***
	(0.0200)	(0.0204)	(0.0155)	(0.0148)	(0.0173)	(0.0169)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes
SOC FE - 1 Digit	Yes	Yes	No	No	No	No
SOC FE - 2 Digit	No	No	Yes	Yes	No	No
SOC FE - 3 Digit	No	No	No	No	Yes	Yes
Location FE - ITL1	No	Yes	No	Yes	No	Yes
Location FE - TTWA	Yes	No	Yes	No	Yes	No
Observations	41095009	41095009	41095009	41095009	37500231	37500231
r2	0.338	0.332	0.383	0.378	0.461	0.459
Standard Errors	clustered	clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.9: Interpersonal skill wage premium and COVID-19 for minimum wage offered

	log_wages_min	log_wages_min	log_wages_min	log_wages_min	log_wages_min	log_wages_min
interpersonal_skill=1	-0.0539**	-0.0546**	-0.0188	-0.0185	-0.0107	-0.0104
	(0.0210)	(0.0219)	(0.0162)	(0.0164)	(.)	(.)
covid	0.0171	0.0177	0.0167*	0.0178	0.0187	0.0200
	(0.0133)	(0.0141)	(0.00869)	(0.00973)	(.)	(.)
post_covid	0.0548***	0.0520***	0.0616***	0.0599***	0.0719	0.0709
	(0.0140)	(0.0135)	(0.0114)	(0.0108)	(.)	(.)
interpersonal_skill=1 × covid	-0.0197	-0.0201	-0.0184*	-0.0185	-0.0172	-0.0172
	(0.0112)	(0.0119)	(0.00987)	(0.0105)	(.)	(.)
interpersonal_skill=1 × post_covid	-0.0227**	-0.0230**	-0.0253***	-0.0256**	-0.0265	-0.0266
	(0.00769)	(0.00826)	(0.00735)	(0.00795)	(.)	(.)
Constant	10.25***	10.25***	10.21***	10.21***	10.20	10.20
	(0.0188)	(0.0193)	(0.0148)	(0.0143)	(.)	(.)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes
SOC FE - 1 Digit	Yes	Yes	No	No	No	No
SOC FE - 2 Digit	No	No	Yes	Yes	No	No
SOC FE - 3 Digit	No	No	No	No	Yes	Yes
Location FE - ITL1	No	Yes	No	Yes	No	Yes
Location FE - TTWA	Yes	No	Yes	No	Yes	No
Observations	41095009	41095009	41095009	41095009	37500231	37500231
r2	0.325	0.319	0.377	0.373	0.455	0.452
Standard Errors	clustered	clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B.5.3 AI Skill

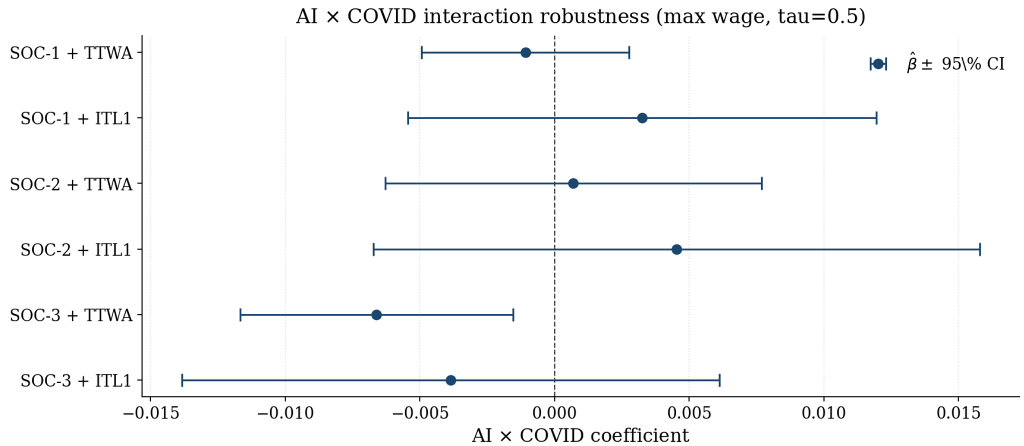

Figure B.9: Wage premium of AI skill during COVID-19 for maximum wage offered

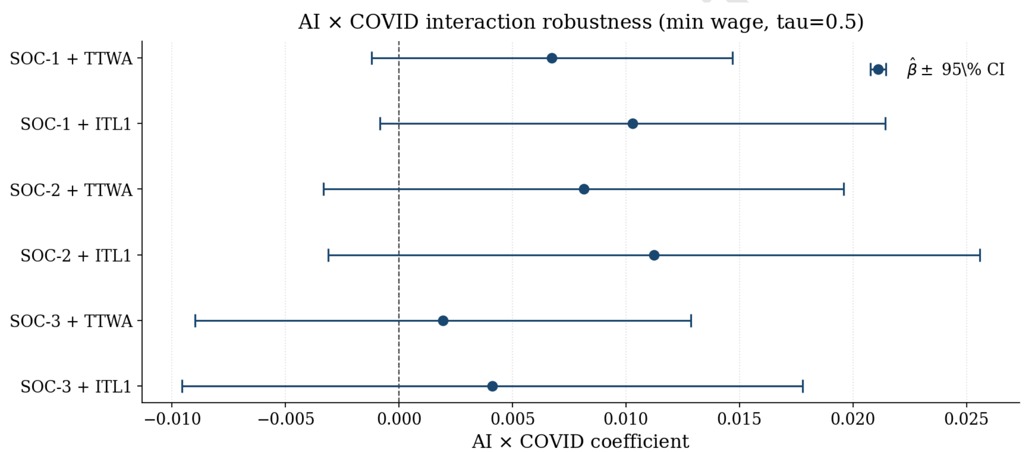

Figure B.10: Wage premium of AI skill during COVID-19 for minimum wage offered

Table B.10: AI skill wage premium and COVID-19 for maximum wage offered

	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
ai_skill=1	0.178***	0.182***	0.112***	0.113***	0.0788***	0.0791***
	(0.0360)	(0.0381)	(0.00927)	(0.00923)	(0.00672)	(0.00668)
covid	0.0432**	0.0435**	0.0420**	0.0429**	0.0482***	0.0496***
	(0.0139)	(0.0153)	(0.0146)	(0.0158)	(0.0120)	(0.0131)
post_covid	0.0559**	0.0525**	0.0559**	0.0533**	0.0677***	0.0662***
	(0.0178)	(0.0176)	(0.0195)	(0.0193)	(0.0155)	(0.0148)
ai_skill=1 × covid	0.00161	0.00631	0.000789	0.00508	-0.00642	-0.00320
	(0.0174)	(0.0191)	(0.0152)	(0.0170)	(0.0115)	(0.0122)
ai_skill=1 × post_covid	-0.0131	-0.0123	-0.00656	-0.00551	-0.00875	-0.00803
	(0.0187)	(0.0192)	(0.0184)	(0.0191)	(0.0159)	(0.0162)
Constant	10.31***	10.31***	10.31***	10.31***	10.31***	10.31***
	(0.0102)	(0.0109)	(0.00908)	(0.00974)	(0.0107)	(0.0111)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes
SOC FE - 1 Digit	Yes	Yes	No	No	No	No
SOC FE - 2 Digit	No	No	Yes	Yes	No	No
SOC FE - 3 Digit	No	No	No	No	Yes	Yes
Location FE - ITL1	No	Yes	No	Yes	No	Yes
Location FE - TTWA	Yes	No	Yes	No	Yes	No
Observations	41095009	41095009	41095009	41095009	37500231	37500231
r2	0.351	0.346	0.388	0.383	0.463	0.461
Standard Errors	clustered	clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.11: AI skill wage premium and COVID-19 for minimum wage offered

	log_wages_min	log_wages_min	log_wages_min	log_wages_min	log_wages_min	log_wages_min
ai_skill=1	0.168***	0.171***	0.102***	0.103***	0.0688***	0.0691***
	(0.0318)	(0.0337)	(0.00741)	(0.00729)	(0.00673)	(0.00675)
covid	-0.00138	-0.00143	-0.00111	-0.000560	0.00301	0.00397
	(0.00908)	(0.0103)	(0.00882)	(0.00969)	(0.00489)	(0.00603)
post_covid	0.0330**	0.0300*	0.0355**	0.0334**	0.0452***	0.0440***
	(0.0136)	(0.0136)	(0.0142)	(0.0140)	(0.0105)	(0.00986)
ai_skill=1 × covid	0.00782	0.0117	0.00658	0.0101	0.000749	0.00332
	(0.0170)	(0.0184)	(0.0155)	(0.0169)	(0.0123)	(0.0129)
ai_skill=1 × post_covid	0.0133	0.0141	0.0186	0.0195	0.0175	0.0181
	(0.0261)	(0.0266)	(0.0253)	(0.0260)	(0.0232)	(0.0236)
Constant	10.18***	10.18***	10.19***	10.19***	10.18***	10.18***
	(0.00791)	(0.00885)	(0.00650)	(0.00724)	(0.00888)	(0.00936)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes
SOC FE - 1 Digit	Yes	Yes	No	No	No	No
SOC FE - 2 Digit	No	No	Yes	Yes	No	No
SOC FE - 3 Digit	No	No	No	No	Yes	Yes
Location FE - ITL1	No	Yes	No	Yes	No	Yes
Location FE - TTWA	Yes	No	Yes	No	Yes	No
Observations	41095009	41095009	41095009	41095009	37500231	37500231
r2	0.339	0.335	0.382	0.379	0.457	0.455
Standard Errors	clustered	clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B.6 Quarter-fixed-effects event study

The three-period (pre, COVID, post) parametrisation of treatment time in Section 4.3 imposes a sharp before-and-after structure on what is in practice a continuous labour-market adjustment over 2016 to 2022. To probe the robustness of the patterns in Tables 4.5 and 4.6, I re-estimate the specification with calendar-quarter fixed effects interacted with each skill, replacing the cubic time control:

$$\log w_i \;=\; \sum_q \beta_q^s \, \mathbf{1}\{\text{quarter}_i = q\} \cdot s_i \;+\; \alpha_{soc2} \;+\; \alpha_{ttwa} \;+\; \alpha_q \;+\; \varepsilon_i, \tag{6}$$

separately for each skill $s \in \{\text{ICT}, \text{AI}, \text{IP}\}$. The quarter-specific premia $\hat{\beta}_q^s$ over the 28 quarters from 2016Q2 to 2022Q4 are plotted in Figure B.11. The ICT premium is consistently positive and shifts only modestly during the COVID window, in line with the small ICT $\times$ COVID coefficient in Table 4.5; the AI premium is positive throughout with no detectable structural break around 2020; and the interpersonal premium hovers near zero throughout. The event study supports the headline finding without imposing the cubic functional form.

C Heterogeneity Analysis

This appendix reports the level and COVID-interaction specifications estimated separately by SOC-1 major group (9 columns) and by ITL-1 region (12 columns). Tables C.1 and C.2 report the four-cell bundle estimates; Tables C.3 and C.4 report the AI premium; Tables C.5 and C.6 report the ICT level and ICT $\times$ COVID interactions; Tables C.7 and C.8 the interpersonal level and interpersonal $\times$ COVID interactions; and Tables C.9 and C.10 the AI level and AI $\times$ COVID interactions. Each column is a separate regression on the corresponding occupation-group or region subsample, with SOC-2 occupation and TTWA location fixed effects absorbed and time controls included. Standard errors are two-way clustered on (year-month, SOC-1) for the regional cuts and on (year-month, ITL-1) for the occupational cuts. The patterns referenced in the main text (Figures 4.1, 4.4, C.3, C.5, and C.7) are read off these tables.

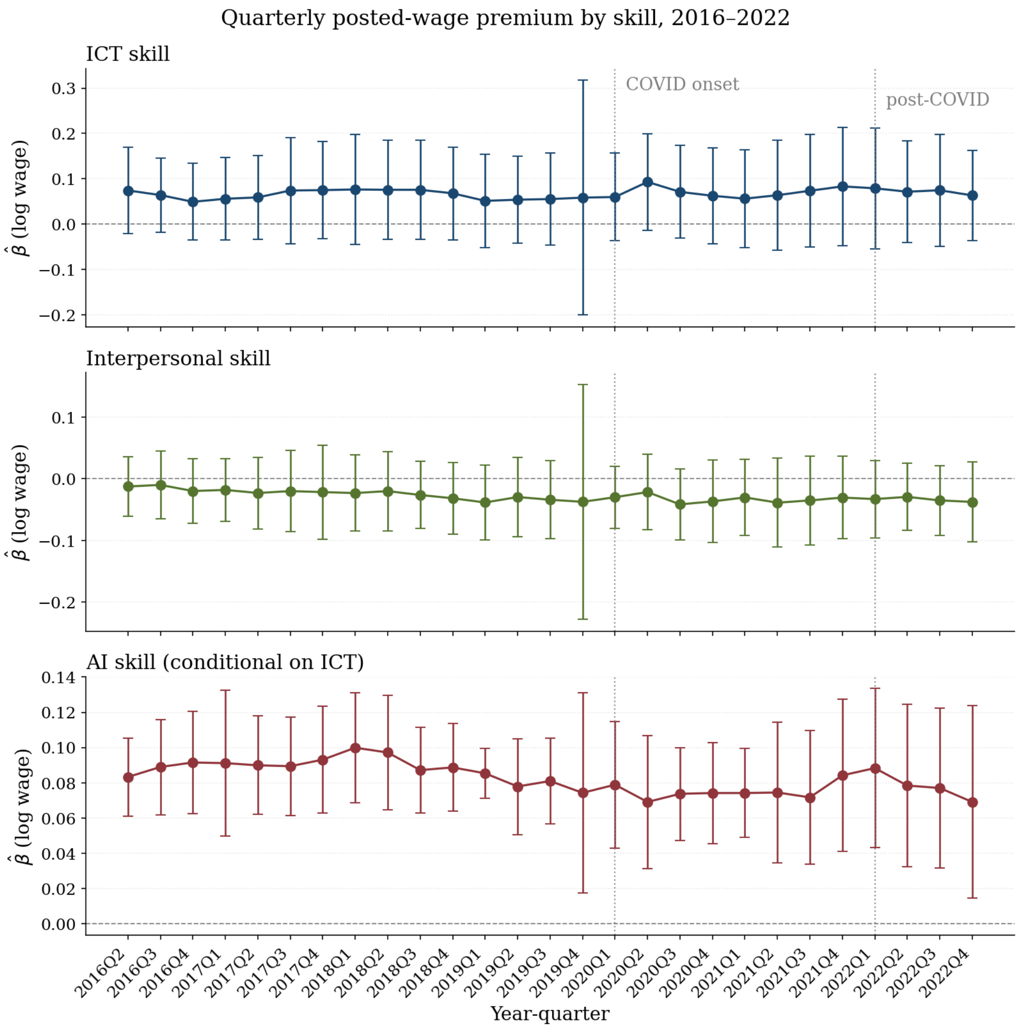

Figure B.11: Quarter-by-quarter wage premia $\hat{\beta}_q^s$ for ICT, AI (at $\tau = 0.50$), and Interpersonal skills, estimated as calendar-quarter fixed effects interacted with each skill (with SOC-2 and TTWA absorbed and quarter-FE included), 2016Q2 to 2022Q4 (28 quarters). Shaded vertical band marks the COVID-19 window (March 2020 to January 2022).

Table C.1: Skill bundle wage premium across occupation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
ict_skill=0 × interpersonal_skill=1	-0.121***	-0.0978***	-0.0932***	-0.0257*	-0.00524	0.0309***	-0.0880**	0.0197***	-0.00915
	(0.00792)	(0.0120)	(0.0186)	(0.0130)	(0.00732)	(0.00859)	(0.0298)	(0.00418)	(0.00514)
ict_skill=1 × interpersonal_skill=0	0.00650	0.0856***	-0.000836	-0.000770	0.0418***	0.117***	0.0181	0.102***	0.0302***
	(0.0112)	(0.0109)	(0.00812)	(0.00464)	(0.0110)	(0.0172)	(0.0243)	(0.0169)	(0.00640)
ict_skill=1 × interpersonal_skill=1	-0.0717***	0.0615***	-0.0868***	-0.0308***	0.0363**	0.0710***	-0.0434	0.131***	0.0379***
	(0.00835)	(0.00530)	(0.0150)	(0.00968)	(0.0120)	(0.00986)	(0.0254)	(0.00987)	(0.00841)
Constant	10.66***	10.57***	10.40***	10.06***	10.25***	9.926***	10.06***	10.07***	9.854***
	(0.0147)	(0.0118)	(0.0220)	(0.0145)	(0.0102)	(0.0230)	(0.0282)	(0.0189)	(0.0437)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3322244	14720195	7629147	3753746	3612013	2056888	2035406	2174905	1790465
r2	0.152	0.198	0.0868	0.156	0.242	0.0507	0.0818	0.0980	0.144
Standard errors	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.2: Skill bundle wage premium across location

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
ict_skill=0 × interpersonal_skill=1	-0.0419	-0.0247	-0.0681^{***}	-0.0398	-0.0256	-0.0339	-0.0445^{*}	-0.0523^{*}	-0.0386	-0.00896	-0.00813	-0.0136
	(0.0330)	(0.0284)	(0.0136)	(0.0239)	(0.0241)	(0.0280)	(0.0209)	(0.0261)	(0.0249)	(0.0244)	(0.0289)	(0.0261)
ict_skill=1 × interpersonal_skill=0	0.0544^{***}	0.0506^{***}	0.0540^{**}	0.0668^{***}	0.0525^{**}	0.108^{***}	0.0726^{**}	0.0502^{***}	0.0587^{***}	0.0835^{***}	0.0682^{***}	0.0585^{***}
	(0.0158)	(0.0121)	(0.0198)	(0.0165)	(0.0161)	(0.0260)	(0.0231)	(0.0144)	(0.0141)	(0.0161)	(0.0160)	(0.0149)
ict_skill=1 × interpersonal_skill=1	0.0253	0.0406^{*}	-0.00318	0.0393	0.0302	0.0555^{*}	0.0591^{*}	0.0209	0.0296	0.0520^{*}	0.0590^{*}	0.0432
	(0.0252)	(0.0211)	(0.0302)	(0.0234)	(0.0265)	(0.0250)	(0.0282)	(0.0236)	(0.0216)	(0.0230)	(0.0259)	(0.0238)
Constant	10.26^{***}	10.22^{***}	10.54^{***}	10.24^{***}	10.24^{***}	10.27^{***}	10.30^{***}	10.32^{***}	10.26^{***}	10.18^{***}	10.25^{***}	10.23^{***}
	(0.0207)	(0.0208)	(0.0207)	(0.0223)	(0.0186)	(0.0196)	(0.0122)	(0.0202)	(0.0158)	(0.0198)	(0.0219)	(0.0186)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3583000	2427748	9952889	922669	4078688	239613	1873937	7182773	3682957	1032212	3442123	2676400
r2	0.350	0.342	0.384	0.337	0.356	0.383	0.380	0.365	0.355	0.355	0.341	0.344
Standard errors	clustered	clustered	clustered	clustered	clustered	clustered		clustered	clustered	clustered	clustered	clustered

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.3: AI skill wage premium across occupation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
ai_skill	0.105***	0.0921***	0.0945***	0.0698***	0.0421***	0.00728	0.0190	0.132***	0.0605***
	(0.00290)	(0.00601)	(0.00881)	(0.00293)	(0.00441)	(0.0167)	(0.0183)	(0.0116)	(0.00838)
Constant	10.60***	10.63***	10.31***	10.02***	10.31***	10.01***	10.01***	10.19***	9.872***
	(0.0162)	(0.0132)	(0.0213)	(0.00765)	(0.00822)	(0.0247)	(0.0223)	(0.0217)	(0.0382)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1812920	8464550	4873778	2559643	1678770	411502	1119984	901062	602854
r2	0.164	0.181	0.0889	0.172	0.143	0.0605	0.101	0.0864	0.119
Standard errors	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: AI skill wage premium across location

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
ai_skill	0.0832***	0.0678***	0.101***	0.0726***	0.0706***	0.0758***	0.0644***	0.0769***	0.0791***	0.0829***	0.0829***	0.0722***
	(0.00825)	(0.00510)	(0.00733)	(0.0134)	(0.00662)	(0.00865)	(0.00924)	(0.00689)	(0.00877)	(0.00806)	(0.00462)	(0.00852)
Constant	10.30***	10.26***	10.59***	10.26***	10.28***	10.33***	10.36***	10.37***	10.31***	10.23***	10.31***	10.27***
	(0.0158)	(0.0133)	(0.0131)	(0.0207)	(0.0190)	(0.0196)	(0.0173)	(0.0173)	(0.0150)	(0.0199)	(0.0154)	(0.0160)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1876356	1280036	5824483	457853	2209319	151422	918668	3936003	1924663	535700	1861367	1449193
r2	0.375	0.370	0.365	0.370	0.379	0.405	0.410	0.385	0.377	0.372	0.362	0.371
Standard errors	clustered	clustered	clustered	clustered	clustered	clustered		clustered	clustered	clustered	clustered	clustered

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.5: ICT skill wage premium and COVID-19 across occupation

	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
ict_skill=1	0.0452***	0.136***	0.00473	-0.0131***	0.0514***	0.0703***	0.0391***	0.130***	0.0420***
	(0.00362)	(0.00726)	(0.00362)	(0.00204)	(0.00800)	(0.0150)	(0.00845)	(0.0139)	(0.00501)
covid	0.0796***	0.00544	0.0928***	0.0520***	0.0403***	0.0201*	0.0617***	0.0158	-0.0313*
	(0.0156)	(0.00985)	(0.0198)	(0.0123)	(0.00877)	(0.0102)	(0.0156)	(0.0141)	(0.0170)
post_covid	0.0938***	0.0158	0.110***	0.0607***	0.0545***	0.0454**	0.0887***	0.0252	-0.00755
	(0.0205)	(0.0194)	(0.0273)	(0.0152)	(0.0123)	(0.0176)	(0.0228)	(0.0198)	(0.0197)
ict_skill=1 × covid	-0.00889	0.0443***	-0.00474	0.0122**	-0.0224***	-0.0641***	0.00133	-0.0322***	0.00988
	(0.00618)	(0.00709)	(0.00533)	(0.00476)	(0.00489)	(0.0108)	(0.00858)	(0.00938)	(0.00702)
ict_skill=1 × post_covid	0.00567	0.0406***	0.000269	0.0259***	-0.0339***	-0.0383**	0.0240***	-0.0400***	-0.00411
	(0.00707)	(0.00697)	(0.0103)	(0.00516)	(0.00380)	(0.0155)	(0.00726)	(0.0110)	(0.00706)
Constant	10.54***	10.48***	10.30***	10.04***	10.23***	9.950***	9.974***	10.07***	9.857***
	(0.0134)	(0.0110)	(0.0185)	(0.00628)	(0.00785)	(0.0266)	(0.0153)	(0.0209)	(0.0446)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3322244	14720195	7629147	3753746	3612013	2056888	2035406	2174905	1790465
r2	0.151	0.197	0.0855	0.157	0.243	0.0523	0.0825	0.0979	0.145
Standard errors	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.6: ICT skill wage premium and COVID-19 across location

	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
ict_skill=1	0.0615	0.0579*	0.0556**	0.0694**	0.0459*	0.0845***	0.0964**	0.0620*	0.0588**	0.0644**	0.0612**	0.0497
	(0.0349)	(0.0298)	(0.0241)	(0.0263)	(0.0232)	(0.0194)	(0.0341)	(0.0303)	(0.0250)	(0.0270)	(0.0242)	(0.0269)
covid	0.0375**	0.0153	0.0420**	0.0272*	0.0333**	0.00169	0.0339*	0.0546**	0.00913	0.0173*	0.0211	0.0349**
	(0.0138)	(0.0148)	(0.0142)	(0.0128)	(0.0141)	(0.0347)	(0.0154)	(0.0216)	(0.0164)	(0.00924)	(0.0186)	(0.0119)
post_covid	0.0651***	0.0204	0.0746***	0.0756***	0.0524**	0.0403	0.0435	0.0609*	0.0242	0.0270	0.0253	0.0661***
	(0.0174)	(0.0212)	(0.0168)	(0.0222)	(0.0187)	(0.0596)	(0.0256)	(0.0272)	(0.0253)	(0.0192)	(0.0220)	(0.0172)
ict_skill=1 × covid	0.00847	0.0135	0.0220**	0.0188	0.0227*	0.0228	0.00395	0.0250*	0.0158	-0.00825	0.0162	0.0239**
	(0.0103)	(0.00981)	(0.00704)	(0.0179)	(0.0119)	(0.0239)	(0.0109)	(0.0111)	(0.0118)	(0.00538)	(0.0136)	(0.00926)
ict_skill=1 × post_covid	0.00135	0.00666	0.0123	0.00928	0.0169*	0.00457	0.00943	0.00262	0.0140	-0.00107	0.00842	0.00566
	(0.00906)	(0.00817)	(0.0139)	(0.0186)	(0.00871)	(0.0332)	(0.0173)	(0.00877)	(0.0159)	(0.00982)	(0.0121)	(0.00651)
Constant	10.22***	10.20***	10.48***	10.21***	10.22***	10.24***	10.26***	10.26***	10.23***	10.17***	10.24***	10.21***
	(0.0232)	(0.0168)	(0.0143)	(0.0171)	(0.0193)	(0.0187)	(0.0145)	(0.0216)	(0.0183)	(0.0306)	(0.0164)	(0.0175)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3583000	2427748	9952889	922669	4078688	239613	1873937	7182773	3682957	1032212	3442123	2676400
r2	0.350	0.342	0.383	0.337	0.356	0.382	0.380	0.365	0.355	0.355	0.342	0.344
Standard errors	clustered	clustered	clustered	clustered	clustered	clustered		clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.7: Interpersonal skill wage premium and COVID-19 across occupation

	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
interpersonal_skill=1	-0.105***	-0.0291***	-0.0885***	-0.0295**	0.0204**	0.0200*	-0.0730*	0.0538***	0.00457
	(0.00820)	(0.00347)	(0.0157)	(0.0105)	(0.00694)	(0.00955)	(0.0344)	(0.00553)	(0.0110)
covid	0.0457***	0.0703***	0.0903***	0.0608***	0.0632***	0.00240	0.0703**	0.0190	-0.0272
	(0.0141)	(0.0115)	(0.0201)	(0.0135)	(0.00875)	(0.0121)	(0.0308)	(0.0163)	(0.0205)
post_covid	0.0733***	0.0728***	0.106***	0.0731***	0.0628***	0.0362*	0.0926**	0.0319	-0.00632
	(0.0211)	(0.0192)	(0.0301)	(0.0152)	(0.0131)	(0.0190)	(0.0392)	(0.0213)	(0.0236)
interpersonal_skill=1 × covid	0.0300**	-0.0457***	-0.000708	-0.000426	-0.0487***	0.00515	-0.00538	-0.0201*	0.0000527
	(0.0122)	(0.00553)	(0.0102)	(0.00474)	(0.00466)	(0.00905)	(0.0241)	(0.0102)	(0.00952)
interpersonal_skill=1 × post_covid	0.0240**	-0.0403***	0.00481	0.00574	-0.0378***	0.00202	0.0122	-0.0246***	-0.00306
	(0.00990)	(0.00662)	(0.0155)	(0.00534)	(0.00539)	(0.00823)	(0.0292)	(0.00792)	(0.0101)
Constant	10.66***	10.59***	10.39***	10.06***	10.24***	9.947***	10.07***	10.09***	9.870***
	(0.0131)	(0.0105)	(0.0189)	(0.0141)	(0.00820)	(0.0235)	(0.0319)	(0.0212)	(0.0460)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3322244	14720195	7629147	3753746	3612013	2056888	2035406	2174905	1790465
r2	0.150	0.180	0.0881	0.157	0.239	0.0472	0.0783	0.0742	0.138
Standard errors	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.8: Interpersonal skill wage premium and COVID-19 across location

	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
interpersonal_skill=1	-0.0167	0.00392	-0.0507**	-0.00781	-0.00720	-0.0381**	-0.0122	-0.0264*	-0.0141	-0.00396	0.00732	0.00362
	(0.0183)	(0.0129)	(0.0186)	(0.0166)	(0.0170)	(0.0149)	(0.0137)	(0.0141)	(0.0155)	(0.0231)	(0.0192)	(0.0175)
covid	0.0644***	0.0459**	0.0691***	0.0696***	0.0681***	0.0162	0.0504**	0.0784***	0.0420***	0.0329	0.0405**	0.0709***
	(0.0156)	(0.0158)	(0.0156)	(0.0153)	(0.0108)	(0.0217)	(0.0160)	(0.0147)	(0.00952)	(0.0183)	(0.0160)	(0.0144)
post_covid	0.0921***	0.0561**	0.0909***	0.120***	0.0844***	0.0580	0.0595**	0.0767**	0.0620***	0.0415	0.0472**	0.0956***
	(0.0203)	(0.0211)	(0.0201)	(0.0265)	(0.0179)	(0.0383)	(0.0214)	(0.0252)	(0.0172)	(0.0259)	(0.0188)	(0.0156)
interpersonal_skill=1 × covid	-0.0265**	-0.0273*	-0.0173	-0.0355***	-0.0252**	-0.000351	-0.0154	-0.0121	-0.0275*	-0.0210	-0.0125	-0.0260
	(0.0106)	(0.0131)	(0.0185)	(0.0103)	(0.0105)	(0.0143)	(0.0171)	(0.0110)	(0.0120)	(0.0210)	(0.0153)	(0.0144)
interpersonal_skill=1 × post_covid	-0.0299**	-0.0371**	-0.0118	-0.0439***	-0.0256***	-0.0123	-0.0139	-0.0175	-0.0331***	-0.0151	-0.0199***	-0.0302**
	(0.00976)	(0.0140)	(0.0202)	(0.00920)	(0.00225)	(0.0166)	(0.0221)	(0.00995)	(0.00796)	(0.0184)	(0.00559)	(0.00914)
Constant	10.27***	10.23***	10.56***	10.25***	10.25***	10.33***	10.32***	10.32***	10.27***	10.21***	10.27***	10.24***
	(0.0188)	(0.0176)	(0.0156)	(0.0244)	(0.0150)	(0.0156)	(0.0173)	(0.0166)	(0.0150)	(0.0186)	(0.0178)	(0.0171)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3583000	2427748	9952889	922669	4078688	239613	1873937	7182773	3682957	1032212	3442123	2676400
r2	0.346	0.337	0.382	0.331	0.353	0.375	0.370	0.361	0.351	0.351	0.337	0.341
Standard errors	clustered	clustered	clustered	clustered	clustered	clustered		clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.9: AI skill wage premium and COVID-19 across occupation

	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
ai_skill=1	0.128^{***}	0.119^{***}	0.108^{***}	0.0699^{***}	0.0660^{***}	0.0421^{***}	0.0621^{**}	0.245^{***}	0.104^{***}
	(0.00391)	(0.00747)	(0.00905)	(0.00471)	(0.0119)	(0.0116)	(0.0220)	(0.0258)	(0.0113)
covid	0.0761^{***}	0.0207^{**}	0.0947^{***}	0.0612^{***}	0.0316^{***}	0.00868	0.0678^{***}	0.00539	-0.0257
	(0.0147)	(0.00937)	(0.0194)	(0.0108)	(0.00814)	(0.0102)	(0.0152)	(0.0138)	(0.0164)
post_covid	0.0983^{***}	0.0278	0.115^{***}	0.0795^{***}	0.0406^{***}	0.0373^{*}	0.106^{***}	0.0156	-0.00751
	(0.0199)	(0.0179)	(0.0280)	(0.0146)	(0.0119)	(0.0174)	(0.0225)	(0.0188)	(0.0192)
ai_skill=1 × covid	-0.0208^{***}	0.0379^{***}	-0.0397^{***}	-0.0105	-0.0124	-0.0425^{**}	-0.0458^{***}	-0.0544^{**}	-0.0330^{**}
	(0.00413)	(0.00666)	(0.00294)	(0.00918)	(0.0102)	(0.0156)	(0.0127)	(0.0222)	(0.0137)
ai_skill=1 × post_covid	-0.0264^{***}	0.0429^{***}	-0.0369^{***}	-0.0135^{*}	-0.0230^{**}	0.0286	-0.0360^{**}	-0.0908^{***}	-0.0406^{**}
	(0.00460)	(0.00687)	(0.00283)	(0.00663)	(0.0101)	(0.0337)	(0.0154)	(0.0146)	(0.0161)
Constant	10.55^{***}	10.54^{***}	10.29^{***}	10.02^{***}	10.25^{***}	9.966^{***}	9.996^{***}	10.11^{***}	9.867^{***}
	(0.0135)	(0.0104)	(0.0183)	(0.00665)	(0.00778)	(0.0241)	(0.0155)	(0.0217)	(0.0444)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3322244	14720195	7629147	3753746	3612013	2056888	2035406	2174905	1790465
r2	0.154	0.190	0.0913	0.159	0.242	0.0477	0.0791	0.0931	0.143
Standard errors	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.10: AI skill wage premium and COVID-19 across location

	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max	log_wages_max
ai_skill=1	0.114^{***}	0.0982^{***}	0.119^{***}	0.103^{***}	0.0823^{***}	0.101^{***}	0.106^{***}	0.0980^{***}	0.104^{***}	0.109^{***}	0.106^{***}	0.0948^{***}
	(0.0135)	(0.00704)	(0.00935)	(0.0118)	(0.00667)	(0.0148)	(0.00967)	(0.00848)	(0.00962)	(0.0124)	(0.00844)	(0.00951)
covid	0.0428^{**}	0.0234	0.0520^{**}	0.0359^{**}	0.0429^{**}	0.0118	0.0371^{**}	0.0645^{**}	0.0164	0.0152	0.0290	0.0467^{***}
	(0.0138)	(0.0138)	(0.0173)	(0.0119)	(0.0133)	(0.0264)	(0.0128)	(0.0207)	(0.0148)	(0.0110)	(0.0157)	(0.0123)
post_covid	0.0660^{***}	0.0258	0.0798^{***}	0.0787^{***}	0.0581^{**}	0.0442	0.0438	0.0611^{*}	0.0310	0.0269	0.0301	0.0692^{***}
	(0.0187)	(0.0211)	(0.0200)	(0.0227)	(0.0190)	(0.0473)	(0.0239)	(0.0271)	(0.0225)	(0.0197)	(0.0196)	(0.0183)
ai_skill=1 × covid	-0.0180	-0.00939	0.00804	0.00386	0.0160	0.0253	-0.0140	0.0167	0.00435	-0.0102	0.00401	0.00254
	(0.0183)	(0.0204)	(0.0121)	(0.0257)	(0.0234)	(0.0146)	(0.0141)	(0.0189)	(0.0165)	(0.0107)	(0.0226)	(0.0192)
ai_skill=1 × post_covid	-0.0180	-0.0244	0.00135	-0.00552	0.0185	0.00523	0.0157	-0.00873	0.00101	0.00528	-0.00439	-0.0143
	(0.0175)	(0.0177)	(0.0143)	(0.0275)	(0.0256)	(0.0356)	(0.0349)	(0.0205)	(0.0225)	(0.0205)	(0.0238)	(0.0226)
Constant	10.24^{***}	10.22^{***}	10.50^{***}	10.23^{***}	10.23^{***}	10.28^{***}	10.29^{***}	10.29^{***}	10.25^{***}	10.19^{***}	10.26^{***}	10.23^{***}
	(0.0121)	(0.0112)	(0.0107)	(0.0191)	(0.0129)	(0.0164)	(0.0158)	(0.0116)	(0.0109)	(0.0202)	(0.0104)	(0.0108)
Time Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3583000	2427748	9952889	922669	4078688	239613	1873937	7182773	3682957	1032212	3442123	2676400
r2	0.351	0.341	0.387	0.336	0.357	0.381	0.375	0.366	0.355	0.356	0.342	0.345
Standard errors	clustered	clustered	clustered	clustered	clustered	clustered		clustered	clustered	clustered	clustered	clustered

Standard Errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C.1 Heterogeneity figures

This subsection collects the regional counterparts of the bundle and AI heterogeneity figures in the main text, and both the occupational and regional cuts of the three COVID-interaction heterogeneity figures. The regional cross-sections are essentially flat in every case, reinforcing the conclusion in the main text that geographic sorting contributes little to the heterogeneity.

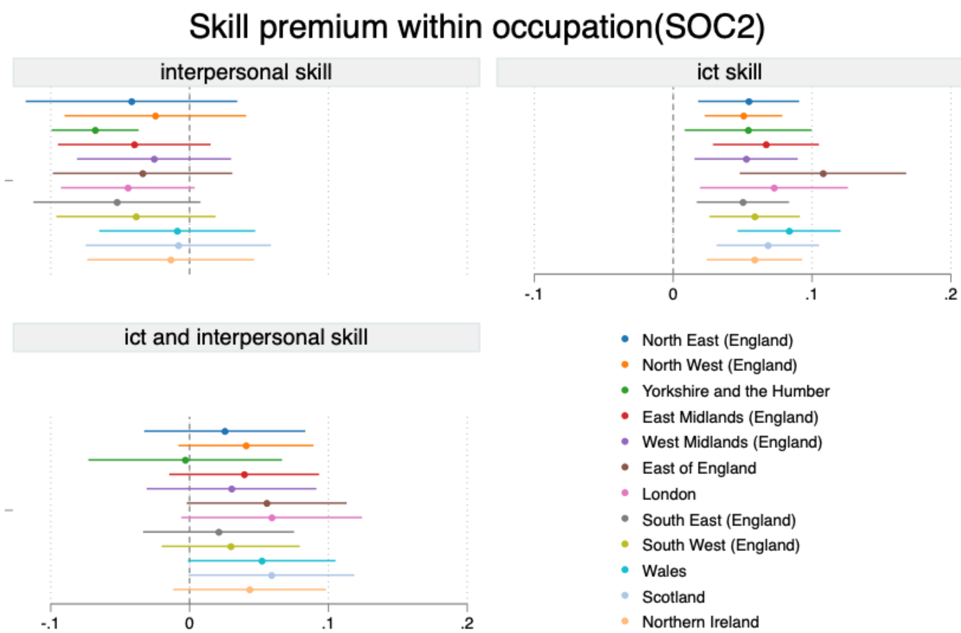

Figure C.1: Wage premium of the skill bundle across ITL-1 regions.

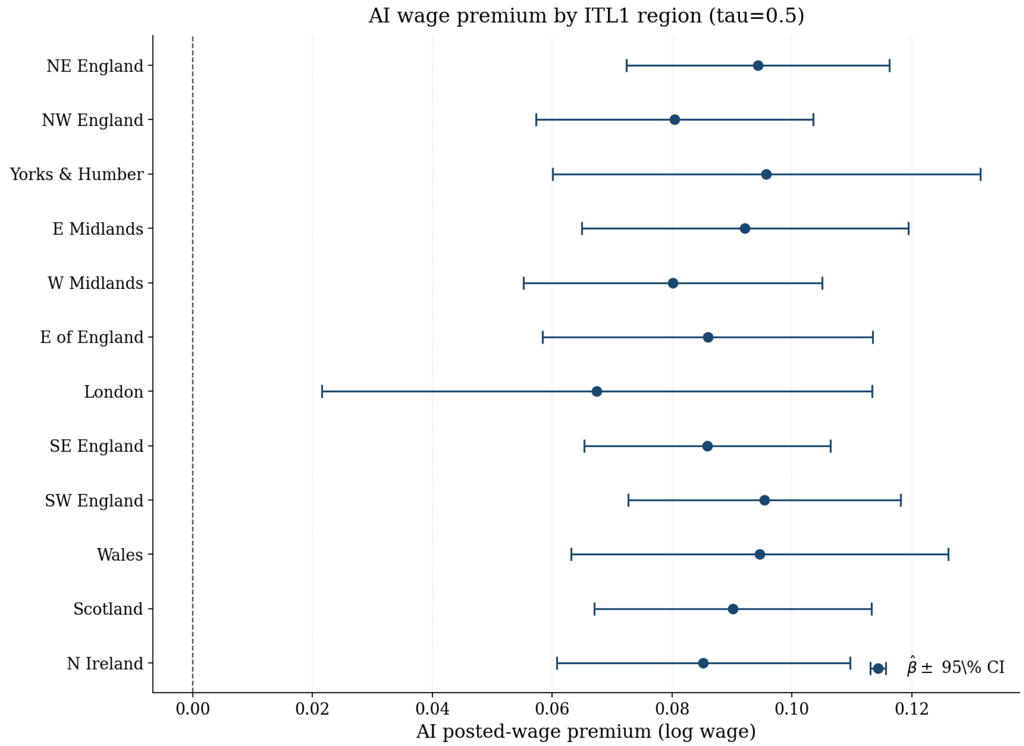

Figure C.2: AI posted-wage premium $\hat{\beta}_{\mathrm{AI}}$ at $\tau = 0.50$ across ITL-1 regions, estimated on the ICT-conditional analytic sample.

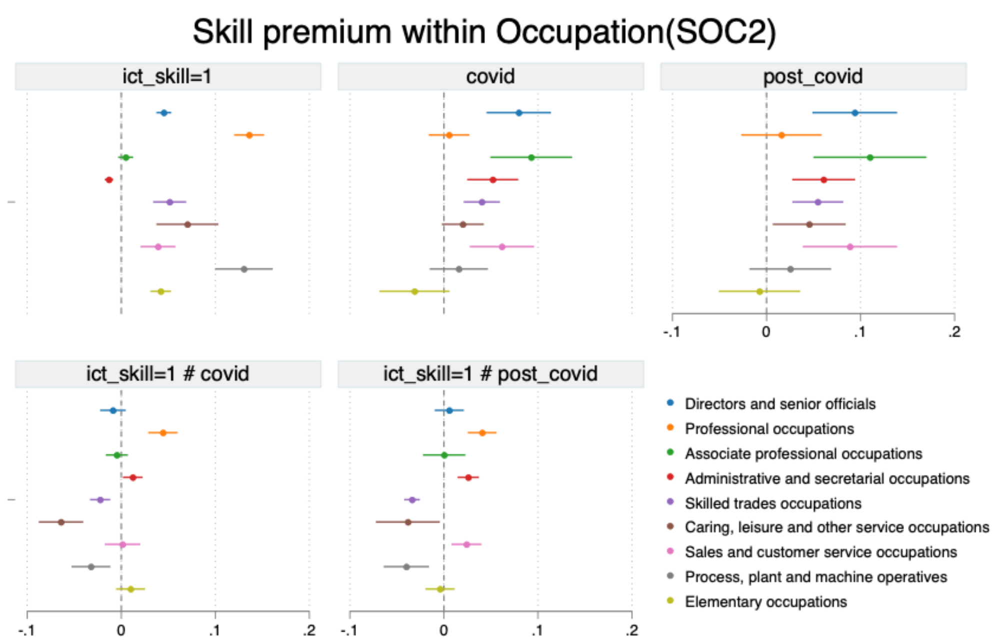

Figure C.3: Wage premium of ICT skills and COVID-19 across SOC-1 major groups.

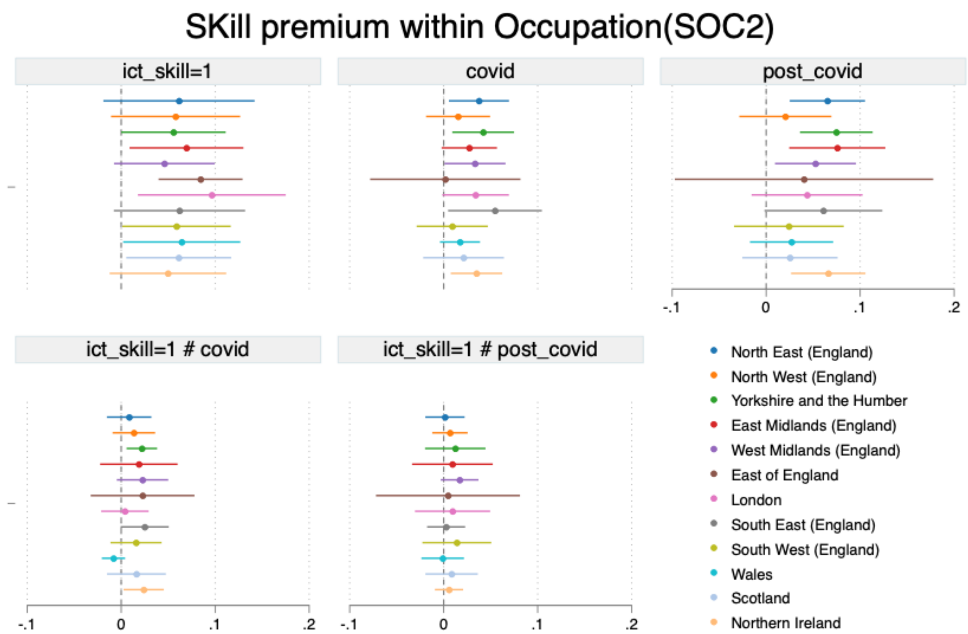

Figure C.4: Wage premium of ICT skills and COVID-19 across ITL-1 regions.

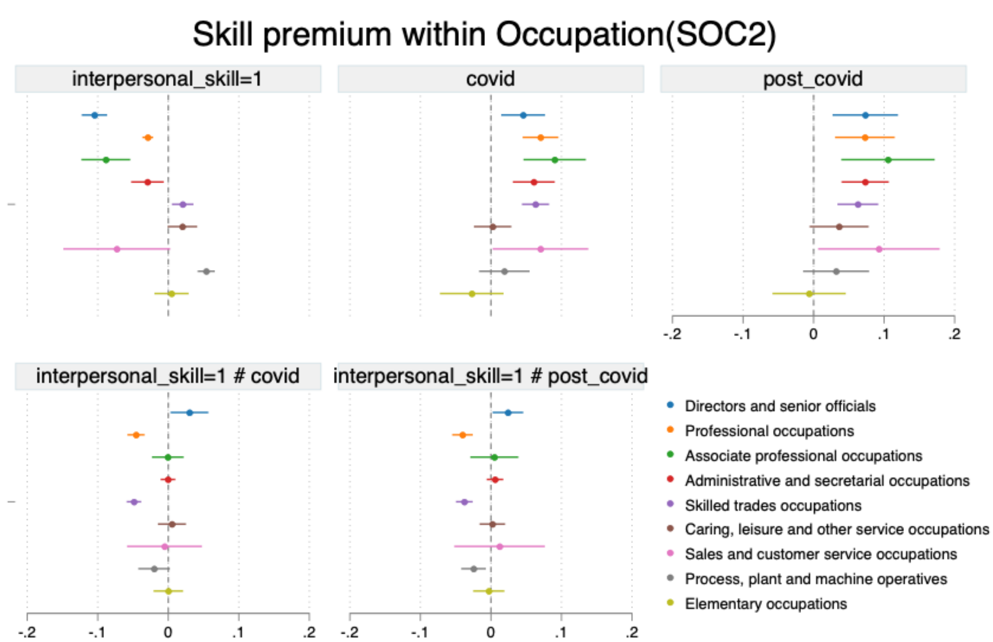

Figure C.5: Wage premium of interpersonal skill and COVID-19 across SOC-1 major groups.

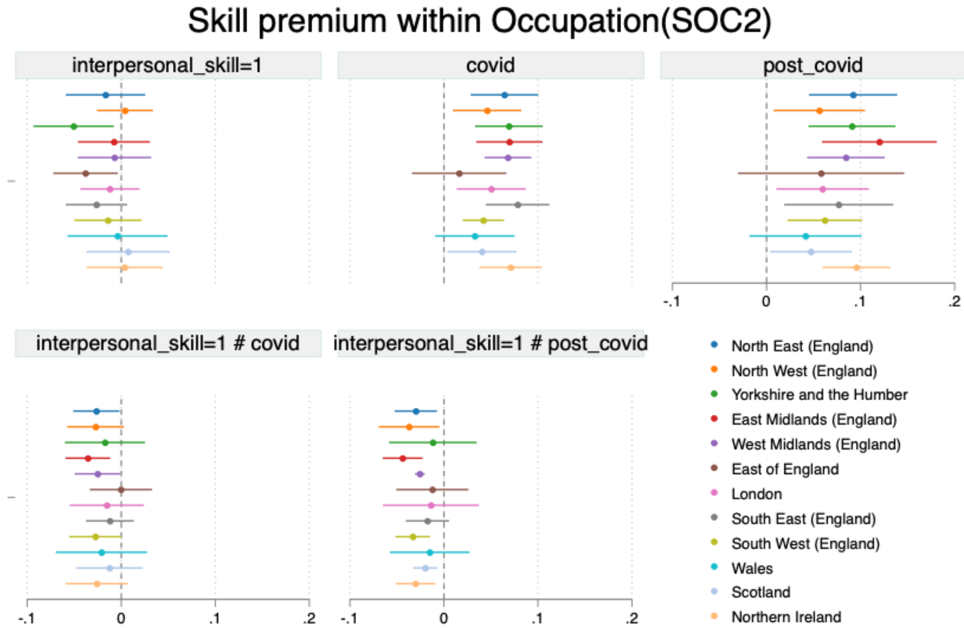

Figure C.6: Wage premium of interpersonal skill and COVID-19 across ITL-1 regions.

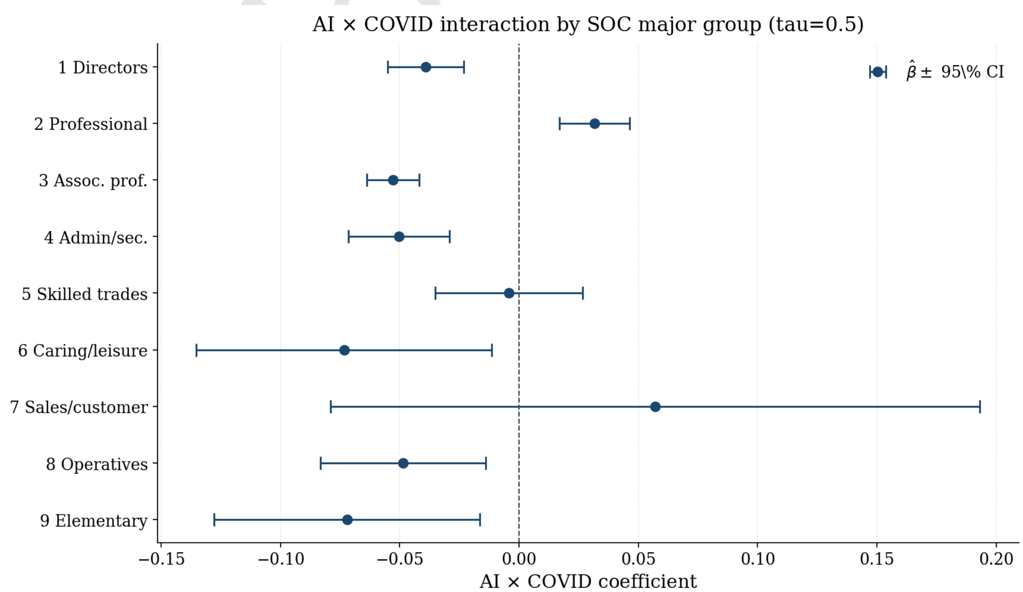

Figure C.7: Wage premium of AI skill and COVID-19 across SOC-1 major groups.

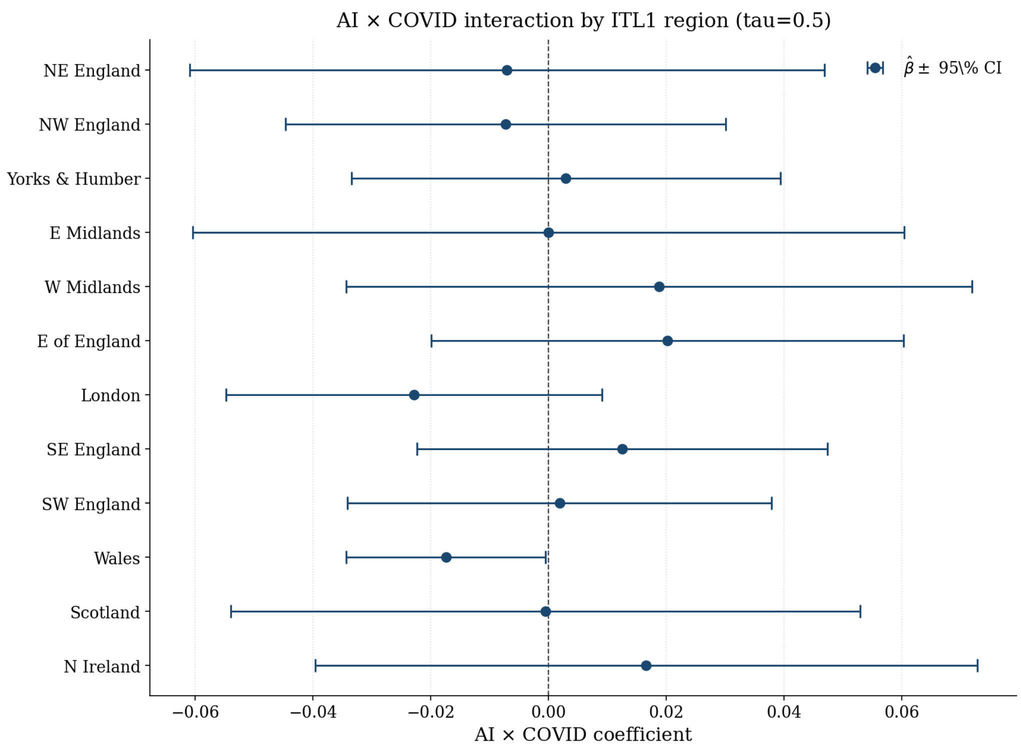

Figure C.8: Wage premium of AI skill and COVID-19 across ITL-1 regions.

The UNU-MERIT WORKING Paper Series

2026-01 *Beyond Monetary Measures: Determinants of Multidimensional Poverty among Vulnerable Older Children and Youth in East Germany. A Perspective from Capability Approach and Capital Theory* by Charlotte Margarete Jost & Zina Nimeh

2026-02 *Sustainable urban rail integration: the case of the city of Maastricht* by Jo Coenen and Luc Soete

2026-03 *Redirect investment to stimulate the economy* by Huub Meijers and Joan Muysken

2026-04 *Airless democracy: Air pollution and voter turnout* by Giulia Rossello, Maria Antonietta Reatini, Gabriele Pinto, and Giorgio Cattani

2026-05 *Beyond size: Unpacking the effects of small-scale de-reservation policy on innovation and performance of manufacturing plants in India* by George Paily, Nanditha Mathew and Carlo Pietrobelli

2026-06 *Deconstructing Quality Infrastructure: A Quantitative Analysis of the Global QI Index* by Ulrich Harmes-Liedtke, Carlo Pietrobelli, Ann-Sara Ramkissoon and Francesco Salustri

2026-07 *The cost of bureaucratic fragmentation: Business tax evasion and revenue mobilization in a low-income country* by Stephan Dietrich, Yannick Markhof and Rose Camille Vincent

2026-08 *Gender and Shock Responsiveness in Social Protection Systems: Lessons from Jordan* by Zina Nimeh, Chris de Neubourg, Míriam Carrera and Nesha Ramful

2026-09 *Automation and Human Capital Investment* by Sidharth Rony

2026-10 *LLM Meets Job Advertisements: Unmasking Skill Premia in the UK* by Sidharth Rony