AI Exposure at Three Levels: Occupations, Tasks and Skills as One Wage Bill

Alessandro Conway

Working Paper AC-WP-2026-04

Version 1.1 (revised September 2026)

Not peer reviewed. Comments welcome.

Abstract

Measures of exposure to generative artificial intelligence are collected at three levels of the labour market, occupations, tasks, and skills, and compared as though they described one object. This paper writes the labour market as one wage bill at three levels and counts what each can identify: 829 observed wages against an identity, against 17,383 task prices that no wage data can contradict, and against 25 skill prices that can. Which quantities in this literature are testable therefore follows from how many statements the analysts of the Occupational Information Network wrote at each level. The leading capability measures are, on inspection, one construct under three names. Converting the Eloundou task ratings to abilities reproduces the Felten ability scores in the right order across AI applications, the Srinivasan automation score is the Eloundou construction re-run on a later release, and a quadratic in automation explains 77 percent of augmentation. The skill layer supports a structure of about twenty-five dimensions, which explains occupational wages at a cross-validated 0.648 in the Network's own named groups, and whose loadings are informative as changes over time but uninformative as levels. On a balanced panel of 693 occupations from 2012 to 2025, nothing in realised wages or employment has moved beyond ordinary volatility since language models were released. No skill group moved by more than 1.51 times its own year-to-year variation, and every deviation from the pre-2020 path begins with the pandemic. If the technology is changing what employers want, the change should appear in job postings before it appears in employment, and the paper specifies that test on the task composition of postings.

Keywords: artificial intelligence, AI exposure, task framework, skill prices, wage bill, O*NET
JEL classification: O33, J24, J23, J31, C43

Suggested citation: Conway, Alessandro (2026). "AI Exposure at Three Levels: Occupations, Tasks and Skills as One Wage Bill." Working Paper AC-WP-2026-04, Version 1.1.

© Alessandro Conway 2026. Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0).

AI Exposure at Three Levels: Occupations, Tasks and Skills as One Wage Bill

Alessandro Conway

First draft May 2026. This version September 2026.

Abstract

Measures of exposure to generative artificial intelligence are collected at three levels of the labour market, occupations, tasks, and skills, and compared as though they described one object. This paper writes the labour market as one wage bill at three levels and counts what each can identify: 829 observed wages against an identity, against 17,383 task prices that no wage data can contradict, and against 25 skill prices that can. Which quantities in this literature are testable therefore follows from how many statements the analysts of the Occupational Information Network wrote at each level. The leading capability measures are, on inspection, one construct under three names. Converting the Eloundou task ratings to abilities reproduces the Felten ability scores in the right order across AI applications, the Srinivasan automation score is the Eloundou construction re-run on a later release, and a quadratic in automation explains 77 percent of augmentation. The skill layer supports a structure of about twenty-five dimensions, which explains occupational wages at a cross-validated 0.648 in the Network's own named groups, and whose loadings are informative as changes over time but uninformative as levels. On a balanced panel of 693 occupations from 2012 to 2025, nothing in realised wages or employment has moved beyond ordinary volatility since language models were released. No skill group moved by more than 1.51 times its own year-to-year variation, and every deviation from the pre-2020 path begins with the pandemic. If the technology is changing what employers want, the change should appear in job postings before it appears in employment, and the paper specifies that test on the task composition of postings.

Keywords: artificial intelligence, AI exposure, task framework, skill prices, wage bill, O*NET.
JEL classification: O33, J24, J23, J31, C43.

1 Introduction

The empirical literature on generative artificial intelligence and work relies on a small number of exposure measures. They have been compared with one another and their disagreements noted, usually as a question of whether capability ratings or observed usage is the more informative. Less attention has gone to the fact that they were collected at different levels of description, so that comparing them at all requires a translation nobody has examined, and less still to the fact that the levels differ in what they can support.

Wages are observed once per occupation, and the 2025 Occupational Employment and Wage

Statistics report 829 of them, carrying a wage bill of 10.70 trillion dollars. The current Network release contains 18,796 task statements, 17,383 of them with the ratings used here, so any set of task values has more unknowns than facts and cannot be contradicted by the data that produce it. Its 161 descriptors reduce to twenty-five groups, so a set of skill values can fail. The paper is organised around that asymmetry. It writes the labour market as one wage bill at three levels and shows at which level each existing measure was collected and how a measure is converted from one level to another. It then checks the conversion where it can be checked, and runs the test the framework allows on the data that exist.

The capability measures, inspected at their source, are one construct under three names. The skill layer supports a structure of about twenty-five dimensions, stated in the Network's own vocabulary, whose loadings are informative as changes over time but not as levels. Realised wages and employment through 2025 show no movement that ordinary volatility does not cover, and every break in the series comes with the pandemic, two years before the technology. The first two results concern the instruments on which this literature depends and hold whatever the technology does next. The third concerns the technology itself: whatever has happened to work, it has not yet changed employment or the prices paid for skills in any measurable way.

Section 2 sets out the three layers as equations and counts what each can identify. Section 3 places Freund and Mann (2026), who build the same nesting in general equilibrium, and sets out what an empirical counterpart needs from them. Section 4 places each measure at its layer and reports what the measures are once inspected, and Section 5 reports what the skill layer supports and how inference on it has to be run. Section 6 tests the Freund and Mann projection on realised prices and employment, and Section 7 concludes and specifies the test on job postings that follows.

2 The three layers

An industry produces output, that output is produced by activities, and those activities draw on human capability, so occupations, tasks, and skills describe one process at three levels of detail. The premise runs from Autor, Levy and Murnane (2003) and Acemoglu and Autor (2011) through the skill literature descending from Katz and Murphy (1992). This section works out the arithmetic of that premise when the only prices the labour market reveals are occupational wages.

Let E_o be employment and w_o the mean wage in occupation o. The first layer is an identity:

$$Y = \sum_o E_o w_o.$$

The identity is a budget: whatever is said at the other two layers has to add up to it, and because it holds by construction it can constrain a description without testing one. Quantities built as shares of Y sum to Y whatever the shares are, so a reconciliation to the national wage bill confirms only that the arithmetic is correct.

The second layer describes each occupation by its tasks. Occupation o performs tasks i in shares β_{oi} that sum to one, and each task has a price p_i:

$$w_o = \sum_i \beta_{oi} p_i.$$

The shares come from the Network's task frequency ratings, published in seven bands from yearly or less to hourly or more and converted to an expected count per year. Frequency is used because importance, the alternative, is compressed: in the median occupation the most important task rates 1.41 times the least important, against a median ratio of 14.78 on frequency. A share built from importance would treat every task in a job as nearly the same part of the work. This is the layer at which the technology has been observed, since every measure of generative AI was built by judging a capability against a description of work. It is also the layer that cannot be tested, with 17,383 prices to recover from 829 wages. The count alone means that the prices are not unique. For the layer to be impossible to contradict, the share matrix must also have full row rank, and it does, because no task identifier in the Network belongs to more than one occupation. Each row of β_{oi} therefore has a support of its own, and some vector of prices reproduces any vector of wages exactly. The 373 statements whose wording repeats across occupations have separate identifiers and separate ratings, and change nothing in that count.

The third layer describes each task by the skills it uses. Task i draws on skills k with intensity γ_{ik}, and each skill has a price π_k:

$$p_i = \sum_k \gamma_{ik} \pi_k, \qquad \text{so that} \qquad w_o = \sum_k s_{ok} \pi_k, \quad s_{ok} = \sum_i \beta_{oi} \gamma_{ik}.$$

The intensities come from the Network's crosswalk from task statements to detailed work activities, which roll up to forty-one generalized activities. These lead in turn, through the published crosswalks from abilities, skills, and knowledge to work activities, to the wider descriptor set. Built that way, 123 of the 161 descriptors vary across the tasks of an occupation. This work began from a construction that set every task's skill mix to its occupation's average, so that all 124,614 occupation by descriptor pairs containing more than one task held a single value, and the skill layer was a fixed linear image of one number per occupation. Rebuilding γ_{ik} from the task crosswalk was the first step. With the rebuilt γ_{ik} and twenty-five skill groups, the regression of w_o on s_{ok} requires twenty-five numbers to reproduce the 742 of the 829 wages that match the Network's content, and the fit can fail. The design has full column rank, with a condition number of 255 on the raw intensities and 16 once the columns are standardised, so the twenty-five coefficients are numerically determined and nothing forces the fit to be good. This is the only one of the three layers at which a claim can fail, and Appendix A describes how the two matrices are built.

The skill grouping used throughout is the third level of the Network's own content model, which gives twenty-five named groups across abilities, skills, knowledge, and work activities. Figure 1 shows the three equations, the two matrices that connect them, and the layer at which each

measure of exposure was collected.

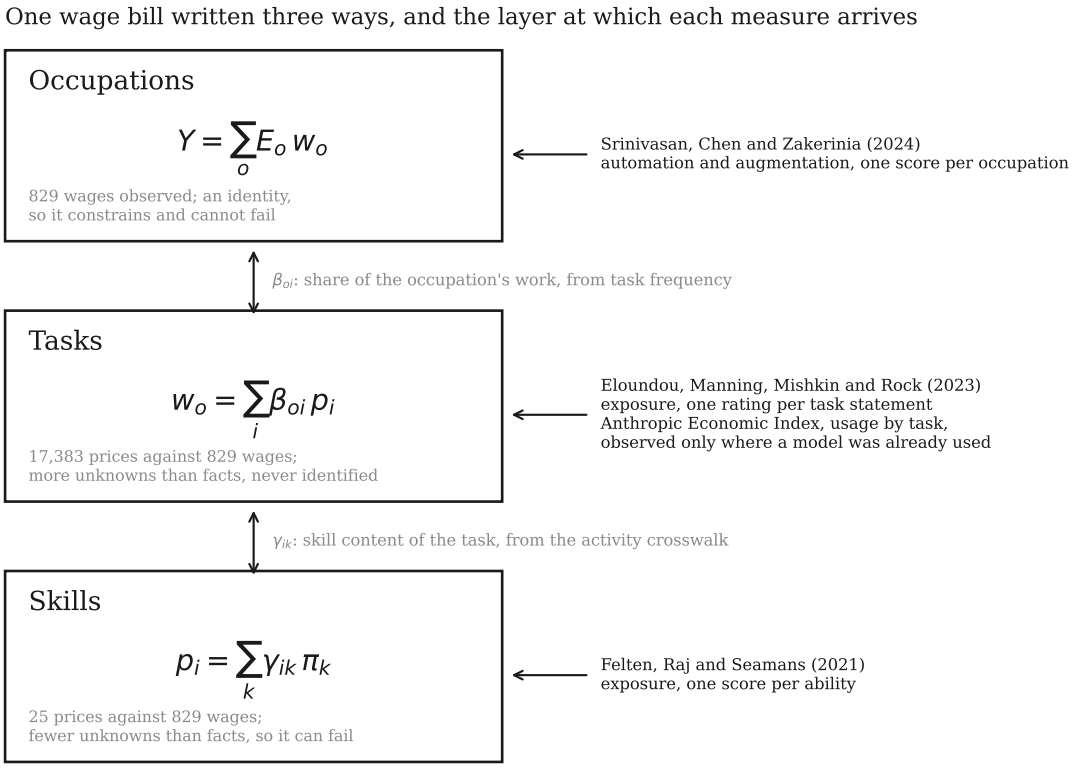

Figure 1: The wage bill at three levels, with each measure placed at the layer where it was collected. The shares β_{oi} convert a task-level quantity to occupations and the intensities γ_{ik} convert it to skills; a quantity collected at occupations can be converted to skills only by inversion.

3 Freund and Mann

Freund and Mann (2026) give the same nesting a general equilibrium form. Occupations bundle tasks, workers hold task-specific skills and sort by comparative advantage, and automating a task raises the weight on the tasks that remain, which is their mechanism of job transformation. Estimating the model on individual worker data, they project that the return to social and non-routine manual skills rises and the return to analytical skills falls. Their task weights come from language model estimates of time shares and their task-specific skill distributions are estimated, so on two of the three layers their construction is the better one.

This paper is built around two features of theirs. They cluster roughly nineteen thousand task statements into thirty-eight categories before estimating, and the clustering is what makes their estimation feasible, which is the counting argument of Section 2 in practice. Their projection, moreover, is what the estimated model implies, and it is never compared with realised prices. An empirical counterpart has to state the projection in a vocabulary that wages can be regressed on, fix the membership of the social and analytical groups before looking at any outcome, and run the test at the one layer where twenty-five numbers set against 829 wages can be wrong. The

framework here does that, and the test in Section 6 is written to their projection.

4 The exposure measures

Felten, Raj and Seamans (2021) enter at the skill layer. They surveyed respondents on the relatedness of ten AI applications to the fifty-two abilities of the Network, combined the resulting matrix with occupational ability requirements, and published the intermediate ability-level scores. Those abilities are fifty-two of the 161 descriptors used here, so the measure joins without propagation, all fifty-two after one renaming between releases. The ten applications come from 2016-era progress benchmarks and include image recognition, real-time video games, and instrumental track recognition, so the headline index measures exposure to artificial intelligence broadly, and the application-by-ability matrix allows any one application to be isolated. For the present technology the language modelling and reading comprehension columns are the relevant ones.

Eloundou, Manning, Mishkin and Rock (2023) enter at the task layer. They rate 19,265 task statements on a four-point rubric separating tasks a language model can substantially accelerate from those it cannot, with separate ratings by human annotators and by GPT-4, of which 18,179 join the task set used here. Their headline index correlates 1.0000 with the machine labels and 0.591 with the human ones, so it is the machine rating, and the distinction matters in Section 5. This is the measure Freund and Mann employ. It is converted to occupations through β_{oi} and to skills through γ_{ik}, and both conversions are used below.

Srinivasan, Chen and Zakerinia (2024) enter at the occupation layer, scoring automation and augmentation for 911 occupations, of which 910 join the Network by name, giving 785 six-digit occupations and covering 681 of the 693 in the panel of Section 6. Both indices are built from the task descriptions of Network release 25.1 classified by GPT-4o on the Eloundou rubric. Job postings are used in their paper to validate the indices and never to construct them, so neither score contains labour market outcomes. An occupation-level score can be converted to the skill layer only by inversion, regressing it on s_{ok}, and that inversion is reported below.

The Anthropic Economic Index also enters at the occupation layer, though it is built from tasks. It classifies conversations with a deployed model against Network tasks and reports the composition of interaction modes for each. Its automation share is the sum of directive and feedback-loop modes and its augmentation share the sum of iteration, validation, and learning, with a filtered residual completing one hundred percent, so the two correlate minus 0.588 by construction. It records usage rather than capability, and only where a model was already in use. Coverage is 20.9 percent of the tasks considered here, the covered tasks are more exposed by 0.236 on a zero to one scale than those omitted, and the most used tenth accounts for three quarters of recorded usage.

4.1 Propagation to abilities

The framework predicts where two measures should agree, and the first check is at abilities, which Felten measured directly and to which the Eloundou ratings can be converted through

γ_{ik} by a route that shares no data with the survey. Propagated exposure correlates 0.667 with reading comprehension, 0.579 with translation, and 0.546 with language modelling, against 0.289 with real-time video games and 0.172 with image recognition. The ordering follows how close each application is to what a language model does, and the video game column serves as a placebo. Since Eloundou's human and GPT-4 raters agree on 65.6 percent of task labels, with numeric scores correlating 0.591, the propagation attains, on the applications that concern it, a correlation of the size at which the source's own raters agree. That comparison sets a correlation across tasks against a correlation across abilities, so it gives a scale for the 0.667 without implying a ceiling that a better propagation could not exceed. The aggregation also preserves variation: averaging eighteen thousand task ratings into twenty-five groups gives a coefficient of variation of 0.292 across a range from 0.045 to 0.513, with mental processes highest, foundational literacy and numeracy close behind, and physical abilities lowest.

4.2 Agreement among the capability measures

The second check is at occupations, where the Srinivasan automation score can be set against the Eloundou ratings, converted to descriptors through γ_{ik} and to occupations through their importance profiles. The two correlate 0.833 because the Srinivasan score applies the Eloundou rubric to Network release 25.1 using GPT-4o, weighting the two exposed categories one and one half and weighting each task by its importance, so the correlation measures reproducibility across data vintages and model versions. Several capability measures in this literature share a rubric or a source of task descriptions, and agreement among them reflects the shared rubric.

Augmentation needs the same care, since Srinivasan and co-authors define it as one minus a Herfindahl index over the shares of exposed and unexposed tasks in an occupation, on the reasoning that a job combining automatable work with work that resists automation gains most. That index is a quadratic in the exposed share, and the automation score is a related summary of the same classification, so a quadratic in automation accounts for 77 percent of the variance in augmentation across occupations.

Inverting the two occupational scores onto the twenty-five skill groups needs no employment weighting, since exposure is not a share of the wage bill, so the effective sample stays at the number of occupations. Skill groups account for occupational exposure with cross-validated fit of 0.712 on automation and 0.415 on augmentation, and the named groups outperform ten principal components on both. On the raw scores, ten of twenty-five groups show an augmentation loading distinguishable from their automation loading, stable across five hundred bootstrap resamples. Removing the quadratic and inverting the residual leaves eleven groups distinguishable, with skill groups accounting for 0.209 of the residual variation, and the two profiles correlate 0.604. Social skills, engineering and technology, arts and humanities, work output, and psychomotor abilities hold on the augmenting side. Business and management falls from a statistic of minus 3.1 to plus 0.4 and mental processes from minus 2.9 to minus 1.2, and so the apparent concentration of automation on cognitive and administrative work largely reflects the construction of the index from an exposure share. Figure 2 shows both profiles.

The two checks establish that propagation assigns a task rating to the right abilities, in the right

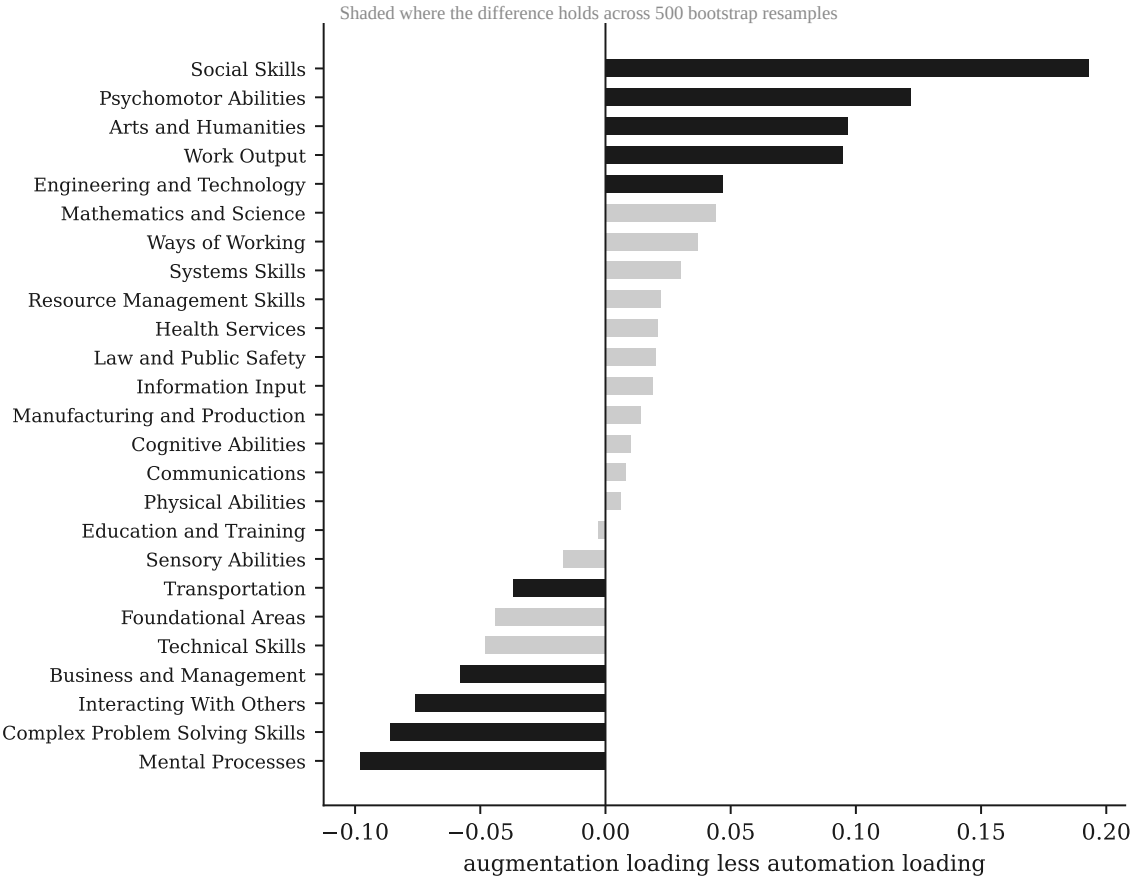

Figure 2: Augmentation loading less automation loading by skill group, from the inversion of the Srinivasan scores onto the twenty-five named groups. Shaded where the difference holds across five hundred bootstrap resamples.

order across applications, and that measures built on one rubric reproduce one another across releases and models. Neither shows that any of the measures predicts an economic outcome, which is the subject of Section 6.

5 The skill layer

The first set of results concerns the Network itself as an instrument. Importance in the Network is recorded on a five-point scale of which respondents use a narrow part, and converting such a rating into a share yields something close to uniform. A descriptor proximity matrix built from occupational similarity places 98.5 percent of its 798,342 pairs above 0.90, with a coefficient of variation of 0.022, and an occupation's profile across the forty-one generalized work activities sits a total variation distance of 0.10 from a uniform profile. The Network's two descriptions of what a job involves, the incumbent survey of work activities and the activity profile implied by the occupation's own tasks, agree at a median rank correlation of 0.33. Valuing 17,383 tasks by allocating each occupation's wage bill in proportion to task importance produces values correlating 0.997 on ranks with dividing that wage bill equally among the occupation's tasks, with eighty-seven of the largest hundred tasks identical under the two procedures. Substituting

frequency lowers the correlation to 0.883. The Network is used because it is free, covers the whole economy, and supplies the vocabulary in which every measure in Section 4 is expressed, and the compression documented here has to be accepted with it.

The second set concerns the skill layer. Fitting occupational wages with descriptor intensities and evaluating on held-out occupations gives 0.554 for a single ordinal measure of required preparation, 0.700 for twenty principal components, 0.648 for the twenty-five named groups, and 0.544 for all 161 descriptors, in 2025 and in the same order in every year from 2019. The named groups give up 0.052 of fit to the best statistical basis of similar size and improve on the full descriptor set, so the wages support a structure of about that dimension, and stating it in the Network's own names costs little. Employment weighting, which a wage bill decomposition requires, gives an effective sample of 121 against 820 occupations in the sense of Kish (1965), the twenty largest occupations carrying a third of employment between them. Skill quantities, employment multiplied by intensity, involve no estimation and are well behaved. The coefficients of the wage regression are wage-skill loadings in the sense of Heckman and Scheinkman (1987), the return an occupation's mean wage shows to a unit of one group's content with the other twenty-four held fixed. They are prices only under the framework's assumption that a unit of a skill earns the same in every occupation. Their levels cannot be used, since eleven of twenty-five groups take a negative loading in some year, and constraining the loadings to be non-negative pins twenty of twenty-five at exactly zero. Differencing leaves the design unchanged, with the same content matrix and the same conditioning, and removes whatever fixed component of an occupation's pay the twenty-five groups do not capture. These are the compensating differentials and rents that persist from one year to the next, and that the levels regression attributes to whichever group happens to covary with them. On the differenced outcome skill content explains 52 percent of the variation in occupational wage changes between 2022 and 2025. The test of Section 6 is run on changes, and its coefficients are read as loadings on wage growth, with the return interpretation resting on the assumption just stated.

Inference on these series has to be built against their own history. Comparing any two periods in these data shows between five and nine of the twenty-five groups apparently changing course when nothing has happened. Conventional standard errors are computed against the precision of a window's average, when the relevant comparison is with the movement of the series between ordinary years. Composition series are smooth enough that a shift worth 0.17 percent of a level per year produces a statistic of 8.2. The procedure adopted throughout estimates the quantity for each of the thirteen year transitions in the panel, removes a linear trend fitted on the transitions up to 2019, and compares the three transitions after 2022 with the other ten by Welch's statistic, the difference of means over its standard error. The statistic is then recomputed for every other way of naming three transitions as treated, the 286 subsets of three among thirteen and separately the eleven consecutive triples. The number reported is the share of those placebo designations at which the statistic is at least as large in the projected direction. That share is a placebo rank against the panel's own history, since the transitions are serially dependent and the placebo windows straddle the pandemic and a classification revision. The consecutive triples are the reference to quote because they share the dependence structure of the treated block.

The last result in this section runs within occupations, and it shapes what the test on postings needs. The Anthropic index records only tasks a deployed model has already been used for, so asking whether exposure predicts usage would select on the outcome. Asking whether exposure predicts the composition of that usage, conditional on the task being used at all, puts the selection on volume and the outcome on mode. On 1,617 tasks across 349 occupations, with occupation means removed, a task's Eloundou exposure predicts the automation share of its observed use at a slope of 0.109 and a standardised value of 3.70 against two thousand within-occupation shuffles. The slope holds when usage is controlled for, and in three of four index releases. Substituting the human rating of the same tasks gives a slope of minus 0.008 and a standardised value of minus 0.38. If the human rating were the machine rating plus classical error, the expected human slope would be the machine slope scaled by the within-occupation covariance of the two ratings over the variance of the human rating, which on these data is 0.025. The observed minus 0.008 sits about one and a half shuffle standard deviations below that, so the human slope alone does not rule out attenuation. The sharper contrast is on the 747 tasks where the ratings differ by more than 0.4, where entering both gives 0.204 to the machine and 0.029 to the human. Both sides of that relationship are language model outputs, so these data cannot distinguish the machine being the more accurate rater from the two sharing method variance, and separating those readings requires a task-level outcome that no model has rated for relevance to AI.

6 Prices and employment

The test is the Freund and Mann projection, stated at the skill layer, on changes, with membership fixed from their vocabulary before any outcome was examined. Social is the Network's social skills and interacting with others groups. Analytical is its foundational areas, complex problem solving, systems skills, mental processes, and mathematics and science groups. Manual is its physical abilities, psychomotor abilities, and work output groups, which they project to rise with social. The quantity under test is the change in what each group earns, relative to its pre-period path, after the release of language models in November 2022, and the projection is that social rises against analytical.

The panel covers 693 occupations from 2012 to 2025, with a wage bill rising from 5.15 to 8.95 trillion dollars. It is assembled from the Occupational Employment and Wage Statistics and mapped across the 2018 revision of the occupational classification using the Network's crosswalk, which raises coverage of the 2018 wage bill from 0.813 to 0.936. Across 2022 to 2025 no skill group moved by more than 1.51 times its own year-to-year volatility. Five of twenty-five exceed two of their own standard errors, against a median of four across the 286 relabellings, and quantities give a placebo rank of 0.455 weighted by headcount. The directional test is not supported on wages and points the other way, the loadings on both social and analytical content having risen against their pre-period paths, with analytical rising further. On employment the social against analytical contrast moves as projected, at a placebo rank of 0.021 against all relabellings and 0.273 against the eleven consecutive triples, on the one specification carrying no pre-existing trend. Manual, which the projection has rising with social, is not supported on either outcome, with detrended statistics of minus 3.11 on wages and minus 1.39 on employment and

ranks of 0.909 and 0.727 against consecutive triples. Figure 3 shows the deviations from the pre-2020 path by year. On wages the residual runs at minus 0.005 in 2020 and reaches minus 0.025 by 2025; on employment the series takes an excursion of minus 0.040 in 2020 and stays positive from 2021. Both breaks precede the technology by two years, and the pandemic moves these series by more than the effect sought.

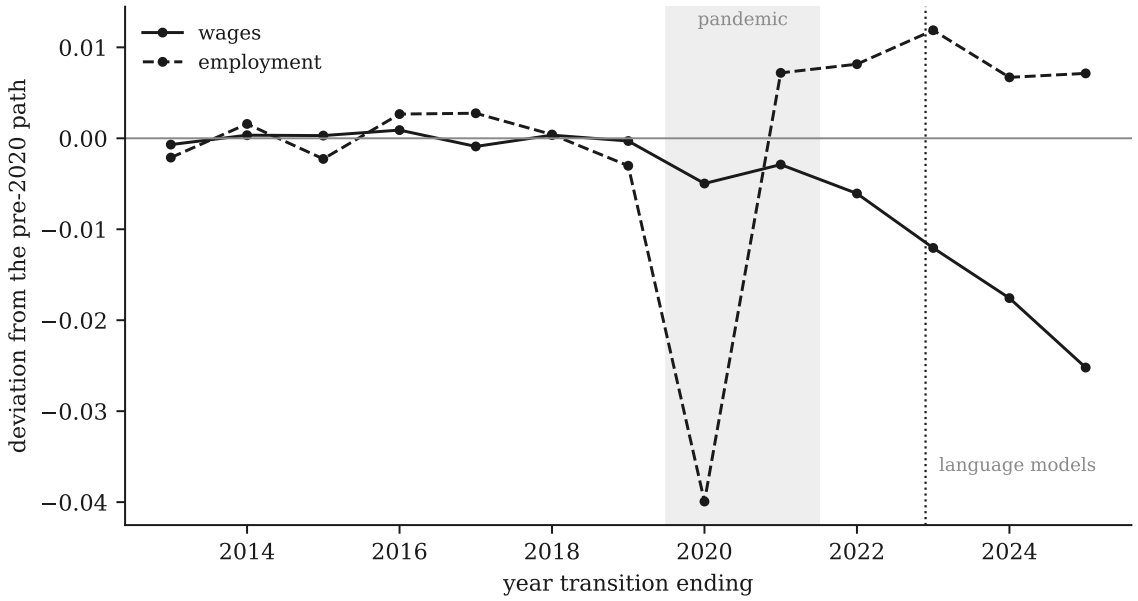

Figure 3: Deviations from the pre-2020 path by year, on wages and on employment, with the pandemic shaded and the release of language models marked.

Employment adjusts slowly, through hiring frictions, notice periods, and sticky wages, so a change in what employers want would appear in job postings some years before it appears in the stock of employment. The result on the stock is therefore evidence only that the technology has not yet had a measurable effect. Three years after the release of language models, the prices paid for skills and the quantities employed of them remain on the paths set after the pandemic, and the only break in the annual series is the pandemic's own. An account of generative AI reshaping the labour market by 2025 would need to explain why the realised wage bill does not yet show it.

7 Conclusion

Measures of exposure to generative artificial intelligence come from three literatures working at three levels of one labour market. This paper writes those levels as one wage bill, shows that only the skill layer can be tested, places each measure at the layer where it was collected, and checks the conversion between layers wherever two measures meet. The conversion from tasks to abilities orders AI applications correctly, and the capability measures prove to be largely a single construct, with augmentation mostly a transformation of automation and a residual that still has skill structure. The skill layer supports about twenty-five dimensions, the Network's own groups lose little to the best statistical basis of that size, and the loadings can be interpreted as

changes, though their levels cannot. Inference has to be built from the rank of the treated years among the panel's own transitions. Within occupations, the machine-rated exposure predicts how a task is used and the human rating of the same tasks does not, so the next test needs an outcome that no model has rated.

Realised wages and employment through 2025 show no movement beyond their ordinary volatility, and every break precedes the technology. Since employment is slow to adjust, the natural next test is the same within-occupation comparison run on the task composition of job postings. Meisenbacher, Nestorov and Norlander (2025) map more than 155 million postings from the National Labor Exchange to Network task identifiers, monthly from 2015 to 2025, so the pandemic and the release of language models are thirty-two observations apart and can be distinguished. Whether a task appears in a posting is an outcome that no model has rated. The estimating equation, its treatment of the exchange's own breaks, its diagnostics for postings written by language models, and the reading of each possible outcome are fixed in Appendix B, and obtaining the series is the work now under way. Should exposed tasks leave postings within occupations while the stock stays unchanged, the gap between what employers ask for and what they employ would be the result. If neither has moved, the null of Section 6 would hold on the fastest-moving series available, and if the stock moves while postings do not, the mechanism is something other than job transformation.

A Construction of the matrices

The task share β_{oi} is built from the Network's task frequency ratings. Each task in an occupation has the share of respondents placing it in each of seven frequency categories, from yearly or less to hourly or more, and the categories are converted to occurrences per year at 1, 6, 24, 100, 250, 750, and 2,000. A task's expected count is the respondent-weighted sum of those rates, and its share is that count divided by the sum over the occupation's tasks, so that shares sum to one within each occupation. The conversion is coarse, and every occurrence of a task counts as the same amount of time whatever the task is.

The skill intensity γ_{ik} starts from the occupation's importance rating for descriptor k and revises it by the task's relevance to that descriptor. Relevance is built in three steps. The Network's crosswalk links each task to one or more detailed work activities, and each link takes an equal share of the task, which is the only weighting the crosswalk supports. The detailed activities roll up to forty-one generalized work activities, giving the task a profile over those activities. A descriptor then takes the share of that profile which its published crosswalk covers, a work activity descriptor taking its own activity and a skill or ability descriptor taking the activities it is linked to. Descriptors with no crosswalk, which is every knowledge element, take a constant relevance and so stay at the occupation average. The intensity is the product of occupation importance and task relevance, renormalised so that each of the four domains keeps the share of the occupation's total importance it had before the revision, which moves weight between descriptors within a domain and never between domains. A shrinkage parameter blends relevance back towards the occupation average. At zero, which is the setting used throughout, 47 percent of the 124,614 occupation by descriptor pairs vary across the occupation's tasks with a median coefficient of

variation of 2.26. At a quarter, 81 percent vary with a median of 0.29, and at one the parameter reproduces the occupation average exactly.

Three operators convert quantities between layers. A task-level score is converted to descriptors as the γ_{ik}-weighted average over the tasks that have a score, which is the aggregation implied by γ_{ik} being a share of the task's labour input, and to a skill group as the unweighted mean over the group's descriptors. A descriptor-level score is converted to an occupation as the importance-weighted average over its descriptors, which is how the Eloundou ratings are compared with the Srinivasan scores in Section 4. The wage regressions of Section 5 and the inversion of Section 4 use occupation-level content, the mean importance over each group's descriptors at the six-digit occupation, so the task-differentiated γ_{ik} enters only where a task-level quantity is carried between layers. The twenty-five groups are the third level of the Network's content model, and the 2025 design of 742 occupations by twenty-five groups and a constant has rank 26.

B The test on job postings

Let y_{oit} be the share of occupation o's active postings in month t that the toolkit of Meisenbacher, Nestorov and Norlander (2025) maps to task i, with the denominator every posting of the occupation that month whether or not it matched any task. Let x_i be the task's Eloundou exposure. The specification is

$$y_{oit} = \alpha_{oi} + \lambda_{ot} + x_i\big(\varphi\,\tau_t + \psi_{m(t)} + \kappa_1 \mathbb{1}[t \geq 2020\text{m3}] + \kappa_2 \mathbb{1}[t \geq t_{2021}] + \theta\,\mathbb{1}[t \geq 2022\text{m11}]\big) + \varepsilon_{oit},$$

where α_{oi} fixes each task's usual share within its occupation, λ_{ot} absorbs everything that moves an occupation's postings as a whole in a month, including its volume, its seasonality, and the exchange's 2021 data warehouse upgrade at t_{2021}. The bracket lets the gradient of posting share on exposure follow a linear pre-period trend φ, a month-of-year seasonal $\psi_{m(t)}$, a level shift at the pandemic, and another at the warehouse upgrade, together with the shift θ at the release of language models, which is the quantity under test. The projection is $\theta < 0$. The unit of observation is the exchange's monthly active job, a posting identifier counted once in each month it is open. A posting an employer reissues under a new identifier is therefore a duplicate the design does not remove, which weights the shares towards employers who post most. The mapping is held fixed by running one version of the toolkit against one Network release for the whole series, the release the exposure ratings join, so the task set is the same in 2015 and 2025. Inference places the break at every month from 2016 to 2019 in turn, re-estimates θ with a treated window of the same length, and reports the rank of the observed θ among those placebo dates, with standard errors clustered by occupation reported beside it.

The outcome is task composition. Mentions of artificial intelligence in a posting are not used, because they follow fashion and an occupation that a language model rates as exposed is also an occupation whose postings are likely to name the technology whether or not the work has

changed. Postings are increasingly written with language models, which the authors read in the rising readability of the text, and a model that rewrites a posting can change how many tasks the toolkit extracts from it without changing what the employer wants. Two diagnostics therefore run alongside the test: the monthly count of matched tasks per posting and the share of postings matching no task, held against the readability series the authors publish. A break in either at November 2022 would indicate a change in boilerplate, and the outcome would then be defined within the posting, as task i's share of the tasks the posting names. The postings cover the jobs that get advertised, which skews towards larger employers, white-collar work, and external recruitment, so any result is about demand at that margin. Access to the underlying series is through the exchange's Research Hub.

References

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: implications for employment and earnings. *Handbook of Labor Economics*, volume 4B.

Autor, D., Levy, F. and Murnane, R. (2003). The skill content of recent technological change: an empirical exploration. *Quarterly Journal of Economics*, 118(4).

Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2023). GPTs are GPTs: an early look at the labor market impact potential of large language models.

Felten, E., 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).

Freund, L. B. and Mann, L. F. (2026). Job transformation, specialization, and the labor market effects of AI. Working paper.

Heckman, J. J. and Scheinkman, J. (1987). The importance of bundling in a Gorman-Lancaster model of earnings. *Review of Economic Studies*, 54(2).

Katz, L. and Murphy, K. (1992). Changes in relative wages, 1963-1987: supply and demand factors. *Quarterly Journal of Economics*, 107(1).

Kish, L. (1965). *Survey Sampling*. Wiley.

Meisenbacher, S., Nestorov, S. and Norlander, P. (2025). Extracting O*NET features from the NLx corpus to build public use aggregate labor market data. Washington Center for Equitable Growth working paper, arXiv:2510.01470.

Srinivasan, S., Chen, W. X. and Zakerinia, S. (2024). Displacement or complementarity? The labor market impact of generative AI. Harvard Business School Working Paper 25-039.