The Convergence of Occupations: Evidence from Online Job Posts

Alexis Antoniades[1]
Georgetown University Qatar

Emmanouil Chatzikonstantinou
Georgetown University Qatar

Mohammed Shahmeer Ahmad
Qatar Computational Research Institute

December 16, 2024

Abstract

We document, in the past decade, a reversal in the trend of increasing wage inequality and declining labor mobility patterns across occupations. This is despite a rapid increase in technological progress and increase in technical skills demanded as evident in online job advertisements. We argue that rapid technological progress over the past decade is transforming occupations and driving them closer together, leading to wage convergence and increased mobility. To study this, we introduce a novel measure of occupational distance based on job postings that allow us to represent each occupation as a vector of skills with corresponding weights that evolve over time and space.

We compute bilateral distances between occupations using similarity measures and find a steady decline in these distances, indicating occupational convergence. We show that this convergence has led to a decline in wage inequality across occupations and increased labor mobility. Furthermore, we find that occupations that have undergone greater transformation—measured as self-distance over time—exhibit higher wage growth and larger reductions in within-occupation inequality.

Keywords: online job vacancies, Euclidean distance, wage inequality, occupations

JEL classification: C55, J31, J62, J63

---

[1] Contact author: Alexis Antoniades, Georgetown University in Qatar, aa658@georgetown.edu; Emmanouil Chatzikonstantinou, Georgetown University Qatar, ec1269@georgetown.edu; Mohammed Shahmeer Ahmad: Qatar Computational Research Institute, mshahmeer99@gmail.com; We thank participants at various conferences for helpful comments received and generous funding by Georgetown University.

1. Introduction

Skill-biased technical change, driven by technological progress favoring certain types of workers and occupations, has been a driving force behind job polarization, declining labor mobility, and rising inequality since the 1980s. However, post-2010, amid further technological advancements driven by innovations related to computing and information technology skills, like big data analytics and artificial intelligence, we observe a convergence in wages across occupations alongside a rise in labor mobility

This trend is illustrated in **Figure 1a**, which plots the coefficient of variation for median wages across occupations, defined at the ONET SOC-4 level, from 1999 to 2022. Wages diverged up until 2008, plateaued thereafter, and have been converging since 2014. A similar pattern emerges when examining the coefficient of variation across occupations at different wage quantiles (**Figure 1b**). Finally, labor mobility across occupations, measured using the longitudinal responses of workers in CPS declined until 2010 but has increased since then, as shown in **Figure 1c**.

Figure 1 – Wage convergence and mobility across occupations in the U.S.

(a) Coefficient of variation for the median wage across SOC4

(b) Coefficient of variation for various wage quantiles

(c) Labor mobility across SOC4

Notes

We argue that technological progress does not only transform tasks and skills performed in each occupation but may also bring occupations closer together, leading to wage convergence and increased occupational mobility. As occupations evolve, tasks and skills become obsolete and disappeare (e.g., typewriting and teller transactions), while new skills (e.g., data analytics and Python) that are highly transferable emerge. Because disappearing skills tend to be occupation-specific, while new skills are adopted across a large number of occupations (as we demonstrate later), the similarity of skills required across occupations has increased[2] . Drawing on the seminal work of Lucas and Prescott (1974), which depicts

[2] The following example illustrates the intuition and contribution of our work. A data scientist can transition seamlessly from research to marketing, agriculture, or banking, whereas someone specialized in teller transactions is confined to banking—and likely not for long. Companies must offer data scientists the prevailing market rate, irrespective of industry or occupation. If a significant set of highly transferable and in-demand skills exists across occupations, we would expect wages to converge and labor mobility to increase as occupations become more homogeneous—effectively reducing the distance between occupation islands, as described by Lucas and Prescott.

occupations as islands with workers distributed across them, our hypothesis is that over the past decade, these islands have moved closer together, effectively reducing the distance between them.

To test this hypothesis, we introduce a novel measure of occupational distance based on online job vacancy data. This measure enables us to assess how similar or dissimilar two occupations are and how this distance evolves over time and across space, using high frequency and current data compared to occupation surveys that tend to be updated irregularly. Specifically, by analyzing hundreds of millions of job vacancies posted in the US between 2014 and 2022 and taking stock of the skills requirements provided in the job description, we represent each occupation as a vector of skills with corresponding weights. We then apply similarity analysis to calculate the distance between these vectors—effectively quantifying the similarity between occupations. By replicating this analysis over time, we can determine whether occupations have converged.

We begin by applying our proposed methodology to measure bilateral distances between U.S. occupations between 2014 and 2022. Using the SOC6 detailed occupation classification, which comprises 1,326 groups, we calculate all bilateral distances between occupation pairs (e.g., *Purchasing Managers, 11-3061* and *Library Technicians, 25-4031*) for each year. This allows us to quantify and observe how these distances evolve over time.

Our first finding is that occupations are indeed converging. The distribution of bilateral distances between 2014 and 2022 has shifted to the left, reflecting a reduction in distances among occupations. Additionally, both the average and median bilateral distances have consistently declined each year within this period. Notably, among aggregate occupation groups, Military (SOC 55) has exhibited the greatest convergence relative to other groups.

Before proceeding to examine the impact of declining bilateral distances on wages and labor mobility, we provide some evidence supporting the validity of the newly introduced skill-based measure of distance. First, as expected, we find that the computed distances between SOC6 occupations within the same SOC2 group (e.g., *Managers, 11*) are smaller on average than distances across SOC2 groups. Second, the smallest distances observed between occupations from different SOC2 groups involve occupations with nearly identical titles (e.g., *Purchasing Managers, 11-3061* and *Wholesale and Retail Buyers, 13-1022*; *Industrial Engineering Technologists and Technicians, 17-3026* and *Industrial Machinery Mechanics, 49-9041*; *Library Technicians, 25-4031* and *Library Assistants, Clerical, 43-4121*). Third, we demonstrate a negative correlation between the computed distances and labor mobility across occupations: labor mobility is lower for occupations further apart. These three tests confirm that our measure of distance is both intuitive and robust and captures information relevant for occupation classification and labor market dynamics.

Between 2014 and 2022, we demonstrate that distances between occupations decreased, wages converged, and labor mobility increased. To move beyond simple correlations, we leverage the richness of the data to test whether occupation pairs that experienced greater convergence also exhibited higher wage convergence and greater increases in labor mobility compared to pairs that did not.

Our regression analysis confirms these relationships. We find a positive and statistically significant coefficient of the estimated bilateral skill-based distance on wage inequality between occupation pairs, indicating that smaller distances are associated with reduced wage disparities. Similarly, we observe a negative and statistically significant coefficient of distance on labor mobility, suggesting that labor mobility increases as occupations become more similar.

Our work makes three key contributions: Theoretically, we provide a mechanism that helps reconcile the empirical evidence of declining across-occupation inequality and rising labor mobility with the predictions of existing occupational models. Methodologically, we introduce a novel skill-based measure of distance, which can also be applied to other groupings, such as industries, firms, and cities. Empirically, we provide evidence of reduction in bilateral distance measures across occupations since 2014 and we confirm that as occupations have converged in their skills requirements, this has led to wage convergence across occupations and an increase in occupational mobility.

The rest of the paper is organized as follows: In Section 2 we present a short review of the literature. The data are presented in Section 3 and the methodology in Section 4. Section 5 presents the empirical findings on the distance between occupations, and Section 6 tests the implications of the distance convergence on wage inequality and labor mobility. Section 7 concludes.

2. Literature Review

Our work contributes to several strands of the literature. Firstly, our paper builds on a growing body of literature that leverages online job postings as a comprehensive snapshot of the labor demand landscape—a strategy pioneered by researchers over the past two decades (Kuhn & Skuterud, 2004; Marinescu, 2017; Deming & Kahn, 2018). In this work, we provide a novel application of online job vacancy data by measuring the similarity between occupations and examining its implications for wage inequality and labor mobility.

Our study explores a new dimension of online job vacancy data, complementing the work of Modestino, Shoag, and Ballance (2020) and Deming and Kahn (2018), which highlights the

importance of such data in understanding trends like upskilling, the evolving task content of occupations, and the growing complementarity between cognitive and computer skills in professional roles.

Secondly, our paper also contributes to the literature examining the dynamics of wages and inequality. Hurst and Kahn (2023) explore the stagnation of median real wages and its implications for income inequality, focusing on challenges in wage measurement, wage cyclicality, and the role of monopsony power in labor markets. Complementing their work, we place greater emphasis on the measurement of occupations and how their evolving nature impacts labor markets and wage dynamics.

By introducing a novel perspective on wage determination, our analysis also complements studies on occupational polarization, such as Autor and Dorn (2013), the effects of occupational mobility (Kampourov and Manowski, 2008), the decline of labor unions (Western and Rosenfeld, 2011), and the changing boundaries of firms and their impact on wages (Bloom et al., 2018). Furthermore, our findings on wage inequality align with recent evidence by Aeppli and Wilmers (2022), who document the stagnation of earnings inequality

Thirdly, our focus on the analysis of the skill content of occupations and especially the transformation of occupations over provide further insights into labor market shifts shaped by technological advancements, particularly through the lens of Skill-Biased Technical Change (SBTC). The SBTC hypothesis suggests that technological progress, especially in information and communication technology, has disproportionately benefited high-skilled workers over low-skilled workers, contributing to wage disparities and shifts in employment patterns.

Early work by Autor, Katz, and Krueger (1998) highlighted the role of computerization in increasing demand for skilled workers, while Acemoglu (2002) provided a theoretical framework to explain how innovations can disproportionately enhance the productivity of skilled workers, thereby driving wage inequality. In contrast, our paper demonstrates that while these dynamics are present, they may be mitigated when technical change transforms occupations in ways that increase occupational mobility. Similarly, Card and DiNardo (2002) questioned the primacy of the SBTC hypothesis as a driver of rising wage inequality, emphasizing the role of institutional factors. Our findings contribute to this ongoing debate by highlighting the interplay between technological change, occupational transformation, and labor market outcomes.

3. Data Sources

Our primary dataset consists of hundreds of millions of online job vacancies (OJVs) posted in the US between January 2014 and March 2022. This data, provided by Lightcast (formerly Burning Glass Technologies), has been widely used by scholars over the past decade. To supplement this, we incorporate wage data from the Bureau of Labor Statistics.

Each day, Lightcast collects OJVs from 40,000 to 50,000 sources, including web boards, company websites, forums, and other directories. These postings typically include details about the job (e.g., title, description, requirements), the ideal candidate (e.g., qualifications, experience, education), the company (e.g., name, industry), and the location (e.g., address). Lightcast processes this unstructured text to create a structured dataset with over 70 parameters for each post, covering aspects such as skill content, company, location, industry, and occupation.

Table 1 – Online Job Vacancies Descriptive Statistics

	Occupations			**Online Job Vacancies** (in thousands, 000s)		**Occupational Characteristics**				
							AI Intensity		**Data Analytics**	
SOC 2	SOC2 Name	SOC4	SOC6	2014 Q1-Q4	2022 Q1	High Nonroutine Cognitive	2014 Q1-Q4	2022 Q1	2014 Q1-Q4	2022 Q1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	23 Total occupation groups (SOC 2)	98	1,326	24,261	12,595	0.45	0.5%	4.6%	2.3%	8.5%
SOC2 occupation group										
11	Management	4	68	2,694	1,522	0.90	0.4%	1.9%	5.1%	10.6%
13	Business and Financial Operations	2	58	1,506	778	0.78	0.4%	1.4%	8.9%	16.3%
15	Computer and Mathematical	2	33	2,528	1,170	1.00	3.3%	15.8%	19.5%	38.7%
17	Architecture and Engineering	3	57	752	333	0.86	0.8%	4.8%	4.6%	10.2%
19	Life, Physical, and Social Science	5	73	253	134	0.95	0.6%	2.3%	9.1%	16.1%
21	Community and Social Service	2	24	283	169	0.79	0.1%	0.2%	2.9%	3.8%
23	Legal	2	12	203	85	0.71	0.1%	0.6%	1.6%	4.5%
25	Educational Instruction and Library	5	94	610	324	0.93	0.1%	0.4%	1.8%	3.0%
27	Arts, Design, Entertainment, Sports, and Media	4	57	619	277	0.72	0.2%	0.7%	1.7%	3.8%
29	Healthcare Practitioners and Technical	3	99	2,381	1,400	0.73	0.1%	0.1%	1.2%	1.7%
31	Healthcare Support	3	24	511	330	0.13	0.0%	0.1%	0.4%	0.7%
33	Protective Service	4	38	265	160	0.53	0.1%	0.6%	1.4%	2.0%
35	Food Preparation and Serving Related	4	28	1,049	622	0.00	0.0%	0.0%	0.2%	0.4%
37	Building and Grounds Cleaning and Maintenance	3	14	323	264	0.00	0.0%	0.0%	0.3%	0.5%
39	Personal Care and Service	8	52	592	294	0.10	0.0%	0.0%	0.2%	0.5%
41	Sales and Related	5	37	3,031	1,248	0.38	0.1%	1.0%	1.4%	2.3%
43	Office and Administrative Support	7	101	2,567	1,360	0.14	0.1%	0.2%	3.7%	7.7%
45	Farming, Fishing, and Forestry	4	22	21	11	0.00	0.0%	0.2%	0.7%	2.6%
47	Construction and Extraction	5	98	282	159	0.08	0.1%	0.4%	0.5%	0.8%
49	Installation, Maintenance, and Repair	4	70	785	430	0.12	0.4%	0.8%	1.1%	1.9%
51	Production	9	159	755	374	0.10	0.1%	0.3%	0.9%	1.9%
53	Transportation and Material Moving	7	86	1,308	634	0.11	0.0%	0.1%	0.2%	0.7%
55	Military Specific	3	22	29	3	-	0.3%	1.7%	0.7%	5.5%

Source: Lightcast, Authors' calculations

Note: This table provides a detailed summary of the characteristics and metrics associated with online job vacancies in the beginning of

the sample period(2014) and the last quarter in the data(2022Q1), encompassing aspects such as the total number of listings and the distribution of vacancies across different occupations. In addition, in the second paper, the share of job vacancies that contain at least one Artificial Intelligence or Data Analytics related skills.

Of particular importance to our analysis are the data on location, occupation, and skill composition. Location information allows us to assign each job posting to a specific geographical area using FIPS codes (county-level identifiers), which can then be aggregated to metropolitan statistical areas (MSAs), commuting zones, states, or other geographic units of interest. Occupation data is provided at the detailed Standard Occupational Classification (SOC 6) level, encompassing 1,326 distinct occupational groups (see **Table 1**, columns 1-6).

Lightcast processes each job posting into a list of keywords and phrases, referred to as skills. Examples of these skills include *Communication Skills*, *Python*, *Bitcoin*, *3D Printing*, *Knowledge of Marijuana*, *Hot Stone Massage*, *Deep Tissue Massage*, *Movement Disorders*, *Plumbing Repairs*, and *Copper Plumbing Installation*. In the first quarter of 2014, the dataset contained 13,989 unique skills, which increased to 14,483 unique skills by the first quarter of 2022.

To illustrate how we utilize the skills data to analyze changes in occupations over time and to support our hypothesis that occupations are converging, we conduct two preliminary exercises. The formal analysis follows in the Section 4.

In the first exercise, we compile a list of skills related to AI and a list related to Data Analytics. These skills are selected based on a combination of consulting reports and academic papers, with the complete lists provided in Appendix A. For each occupational group (SOC 2), we count the number of job vacancies that mention each specific skill. This process is repeated for all skills available during the analyzed period, generating a skill vector for each SOC 2 group along with a corresponding measure of frequency, referred to as document frequency (df).

To calculate the AI intensity of job vacancies in a particular SOC 2 group, we sum the df of all AI-related skills identified on our list and divide it by the total df of all skills (AI and non-AI). We apply the same methodology to compute the Data Analytics intensity for each SOC 2 group.

In **Table 1**, Columns 8 and 9 report the AI intensity for 2014 and 2022Q1, respectively, while Columns 10 and 11 present the Data Analytics intensity. Additionally, we include the share of SOC6 occupations within each SOC2 group classified as High Nonroutine Cognitive (High NR-C). This classification is based on Acemoglu and Autor (2021), which utilizes approximately 250 occupational measures of tasks, skills, and capabilities from O*NET. The importance of these measures is assessed at the 6-digit occupation level, and the

Occupational Employment Statistics (OES) is used to aggregate these classifications to match the 6-digit SOC6 framework in the Lightcast data.

The first row of **Table 1** presents aggregated data across all occupations. The dataset includes job vacancies allocated to 1,324 occupations at the SOC6 level, which can be aggregated into 98 SOC3 or 23 SOC2 groups. In 2014, the dataset contains 24,261 observations, while in the first quarter of 2022, it includes 12,595. Among these 1,324 occupations, 45% are classified as High Nonroutine Cognitive (High NR-C). Our calculations show that AI intensity increased from 0.5% in 2014 to 4.5% in 2022, while Data Analytics intensity rose from 2.3% to 8.5% over the same period.

In the subsequent rows, the data are broken down by SOC2 categories, of which there are 23. Across all occupations, we observe a significant increase in AI and Data Analytics intensities between 2014 and 2022, though the dynamics vary by occupation. Notably, occupational groups with a higher share of High NR-C roles in 2014 had a larger share of jobs requiring AI or Data Analytics skills. However, the growth in demand for these skills over time does not appear strongly correlated with the High NR-C classification.

For instance, in Community and Social Services (SOC2 21) and Educational Instruction and Library (SOC2 25), 79% and 93% of SOC6 occupations, respectively, are classified as High NR-C. Yet, AI intensity in these groups increased only marginally, from 0.1% in 2014 to 0.2% and 0.4%, respectively, by 2022. In contrast, in Protective Services (SOC2 33) and Sales and Related (SOC2 41), where the share of High NR-C roles is lower at 53% and 38%, the demand for AI skills grew more substantially—from 0.1% in 2014 to 0.6% in 2022 for Protective Services and 1.0% for Sales and Related. A notable case is the military (SOC2 55), where AI intensity increased from 0.3% in 2014 to 1.7% in 2022, while Data Analytics intensity rose from 0.7% to 5.5%.

In our second exercise, we move beyond a curated list of AI and Data Analytics skills to investigate broader trends. Specifically, we examine which skills, out of over 15,000 in the dataset, experienced significant changes between 2014 and 2022. We classify these skills into three categories:

1. **Common Skills:** Skills that existed in both 2014 and 2022 and experienced the highest growth.
2. **New Skills:** Skills that did not exist in 2014 but gained significant prominence in 2022 (based on high document frequency).
3. **Disappearing Skills:** Skills that were important in 2014 (based on a document frequency threshold) but saw a decrease of 85% or more by 2022.

Figure 1 – Growing Skills and Fading Skills, 2014 – 2022

(a) Common: Skills with the highest growth in demand

(b) New: Skills in high demand, 2022 (not present in 2014)

(c) Disappearing: Skills with 85% or more drop in demand, sorted by 2014 importance

Source: Lightcast, Authors' calculations

Note: Each word cloud visually summarizes the key terms and technologies that are prevalent in the Lightcast data. The size of each term corresponds to its frequency or importance, with the first panel using only skills important in both 2014 and in 2022, the second panel concentrating in new skills, and the last panel using disappearing skills: those skills with 85% or more drop in demand .

To visualize these trends, we present WordClouds for each of these categories in Figures 1A, 1B, and 1C. To further underscore our paper's motivation, we repeat this analysis for three specific occupational groups: Business Operation Specialists (SOC4 13-10), Engineers (SOC4 17-20), and Lawyers, Judges, and Related Workers (SOC4 23-10). The results for these groups are displayed in **Figure 2**.

Two key findings emerge from our analysis: First, within just five years, numerous new skills have emerged, most of which are technology-related. This trend is clear when examining the fastest-growing existing skills between 2014 and 2019, as well as the new skills that ranked highest in employer demand in 2019. These patterns hold true whether we analyze data across all occupations (**Figure 1**) or focus on specific occupational groups (**Figure 2**).

Second, Figure 2 reveals that the skills with the highest growth in demand tend to be common across the three distinct occupational groups. In contrast, the skills that are disappearing vary significantly between these groups. This indicates that while occupations evolve over time by shedding older, occupation-specific skills, they increasingly adopt new, occupation-agnostic skills.

This dynamic of skill entry and exit, as demonstrated in the word cloud analysis, suggests that occupations are converging. To formalize this observation, we introduce a measure of distance in the next section.

Figure 2 – Growing Skills and Fading Skills for Selected SOC4 Groups, 2014 – 2022

4 Methodology

We introduce a novel method to measure the distance between groups of workers based on the skill content of jobs, as derived from online job vacancies. Unlike the work done using tasks - which are invariant to space and do not change frequently as they rely on surveys-- to measure similarity, our distance metrics can be both - and space-specific and can be customized to any particular group (corpus) of interest.

For each group of interest g, we compile all online job postings related to that group at time t. A group can represent an occupation, city, industry, firm, or any other defined set. This generates a corpus C_{gt} of online job postings for group g at time t.

Next, for each skill s within the total set of skills available during that period, we calculate the number of documents in the corpus C_{gt} that mention the skill. This count is referred to as the document frequency (df_{st}). We then compute the intensity of each skill as the share of df_{st} relative to the sum of df_{st} values for all skills in the corpus. This process reduces the corpus into a skill vector with corresponding shares: $C_{gt}=(c_1,c_2,\ldots,c_S)$.

To calculate the distance between two corpora, we represent them as vectors $A=(a_1,a_2,\ldots,a_S)$ and $B=(b_1,b_2,\ldots,b_S)$. We then compute the Euclidean distance between these vectors, defined as:

$$Distance(A,B) = \sqrt{\sum_{i}^{S}(a_s - b_s)^2}$$

Given that there are over 14,000 skills, most of which have values of zero, we limit our analysis to the top 200 skills in corpus AAA to reduce computational complexity. Specifically, both vectors are sorted based on their aia_iai values, ranked from largest to smallest, and the weights of the top 200 skills are compared across the two vectors. In the next section, we demonstrate that the results are robust to varying this cutoff, whether smaller or larger. For further robustness, we replicate the analysis using Cosine similarity as an alternative metric.

When the two corpora correspond to different occupations at time ttt, the Euclidean distance measures the difference in the skill mix of the two occupations. By repeating this calculation for a later period t+Tt+Tt+T, we can assess how the distance between the two occupations has changed over time, thereby determining whether convergence across occupations has occurred.

Additionally, by comparing the distance of an occupation across two points in time (i.e., corpus C_{gt} and $C_{g(t+T)}$), we can measure the extent of transformation within that occupation over time. A larger distance indicates greater transformation. To distinguish this self-

comparison from the bilateral distance between two groups, we refer to this metric as *self-distance*.

This methodology is flexible and can be applied beyond occupations. It enables us to compute skill-based distance measures for various groups, such as cities, counties, industries, firms, or commuting zones, providing a versatile framework for analyzing transformations across different entities.

5. Analysis

In this section, we illustrate the application of the newly introduced similarity measure to compute skill-based distances between occupations. The analysis begins with the calculation of bilateral distances between occupation pairs, providing a systematic way to assess similarities in their skill requirements. These distances offer a quantitative foundation for understanding the relationships among occupations.

To explore how these relationships evolve, we analyze changes in occupational distances over time. This dynamic perspective reveals patterns of skill convergence or divergence, shedding light on broader trends in the labor market.

Following the computation of these distances, we present validity checks to ensure the robustness and reliability of the results. We also discuss how by representing occupations as nodes and their skill-based distances as edges, we can construct networks that reveal structural patterns within the occupational space.

5.1 Bilateral distances over time across SOC6

We begin the analysis by computing bilateral measures of distance between occupations at the SOC6 level over time. For each year between 2014 and 2022, we use the OJV data to summarize the skill landscape of each occupation during the first quarter (Q1). Our focus on Q1 is driven by data availability, as 2022 data exist only for that quarter. Using our proposed similarity measure, we calculate the bilateral distances between all SOC6 occupations for each year.

To illustrate how occupational distances evolve, **Figure 3a** presents kernel density plots of the computed bilateral distances for the years 2014, 2018, and 2022. The results show a clear leftward shift in the distribution over time, while the shape of the distribution remains largely unchanged. This indicates that occupations are converging—distances between them are decreasing, and they are becoming more similar in terms of skill content.

We further demonstrate convergence in **Figure 3b**, which reports the average and median bilateral distances across all SOC6 occupation pairs for each year. Both measures consistently decline over time, reinforcing the finding that occupational distances are narrowing. Importantly, this result is robust to the choice of distance metric: when the Cosine Similarity method is applied instead of the Euclidean measure, the conclusions remain unchanged (see Appendix B1).

These findings suggest a clear trend of skill convergence across occupations, providing new insights into how occupational roles have evolved in terms of their skill requirements over the past decade.

We further explore whether convergence depends on the nature of tasks performed within an SOC6 occupation. To do this, we classify occupations into **High Nonroutine Cognitive (High NR-C)** and **Low Nonroutine Cognitive (Low NR-C)** groups, following the framework of Acemoglu and Autor (2021). For each year, we compute and report the average bilateral distances for three occupational pairings: (i) **High NR-C – High NR-C**, (ii) **Low NR-C – Low NR-C**, and (iii) **High NR-C – Low NR-C** (or equivalently **Low NR-C – High NR-C**).

The results, presented in **Figure 3c**, reveal that **Low NR-C** occupations are more distant from both other **Low NR-C** occupations and **High NR-C** occupations. In contrast, **High NR-C** occupations are notably closer to other **High NR-C** occupations, reflecting greater skill similarity within this group. Nonetheless, bilateral distances are declining across all categories over time, reinforcing the trend of overall convergence in skill content among occupations.

It is also of interest to identify which occupational groups have experienced the most significant changes over time. To do this, we calculate the average bilateral distances d_{ij} for all i across all j, where i and j represent SOC6 occupations, in both 2014 and 2022. We then measure the change in the average distances when grouped by SOC2 group to assess convergence.

The results, presented in **Figure 4**, reveal that all occupational groups at the SOC2 level have converged over time. However, the magnitude of convergence varies across groups. *Military* exhibits the most substantial convergence, while *Social Services* shows the least.

5.2 Validity of Computed Distances and Robustness

Before we proceed any further, we want to ensure that this newly introduced skill-based measure of distance does indeed capture similarity. We show that it does using three different approaches.

Figure 3 - Bilateral distances across SOC6 occupations

a. Occupations' bilateral distances distribution for 2014, 2018, and 2022

b. Average and Median Euclidean Distance Across All Occupations by Year

c. Median Euclidean Distance by Year and Occupation Type (High vs Low Nonroutine Cognitive)

Source: Lightcast, Authors' calculations

Note: *Figure 3 provides an analysis of the evolving relationships between different occupations over time, as measured by bilateral distances. Panel (a) depicts the distribution of these distances among SOC6 occupations across three selected years (2014, 2018 and 2022), highlighting the change in occupation similarity. Panel (b) aggregates this data to present the overall trends in average and median Euclidean distances. Lastly, panel (c) breaks down the median Euclidean distances into occupation types, categorized by the degree of nonroutine cognitive skills required, illustrating the distinctive patterns of occupational similarity over time.*

Figure 4 –Change in Average Distance Between 2014 and 2022 by Occupation Group

Source: Lightcast, Authors' calculations

Note: This *Figure portrays the average change in Euclidean distances for occupations. A negative change denotes a convergence in occupational characteristics or requirements to all other occupations on average, suggesting that certain jobs have grown more akin in their skillsets or functions. Notably, low skill occupations exhibit the largest decrease in distance. This bar chart describes the trajectory of occupational similarity, reflecting how the nature of work has become more integrated across these fields*

First, we expect that occupations within the same SOC2 group (e.g. Management) should be more similar (shorter distance) compared to occupations across SOC2 groups. We show this in **Figure 5** below, where we plot the average bilateral distance d_{ij} for SOC6 occupations *i,j* within the same SOC2 group (*Within*) and the average distance when *i,j* come from different SOC2 groups (*Across*). The figure demonstrates that computed *Within* distances are always smaller than *Across*, as expected, and that both distances decrease over time.

Figure 5 –Occupation Distance, Within and Across SOC2, 2014-2022

Next, we rank the bilateral distances d_{ij} across SOC2 groups from smallest (most similar) to largest (most dissimilar) and present the top 25 occupational pairs in **Table 2**. The results reveal that the closest pairs often have nearly identical names or highly overlapping job roles.

For example, the smallest distance is observed between **Purchasing Managers** (11-3061; SOC2 = *Management*) and **Wholesale and Retail Buyers** (13-1022; SOC2 = *Business and Financial Operations*), highlighting their strong skill alignment. Similarly, **Industrial Engineering Technologists and Technicians** (17-3026; SOC2 = *Architecture and Engineering*) are most closely related to **Industrial Machinery Mechanics** (49-9041; SOC2 = *Installation, Maintenance, and Repair*). Another notable pair is **Library Technicians** (25-4031; SOC2 = *Educational Instruction and Library*) and **Library Assistants, Clerical** (43-4121; SOC2 = *Office and Administrative Support*), which reflect highly similar roles despite belonging to different SOC2 categories.

The ability of our methodology to identify and link such closely related occupations—based solely on skill information extracted from job advertisements—demonstrates the robustness and accuracy of our approach. The fact that these relationships emerge naturally from the data further validates the reasonableness and practical utility of the computed bilateral distances.

Table 2 – Top 25 SOC6 pairs with the lowest distance, across SOC2 groups, 2014

Rank	Occ 1	Occ 2	Occupation 1 Name	Occupation 2 Name
1	11-3061	13-1022	Purchasing Managers	Wholesale and Retail Buyers, Except Farm Products
2	17-3026	49-9041	Industrial Engineering Technologists and Technicians	Industrial Machinery Mechanics
3	25-4031	43-4121	Library Technicians	Library Assistants, Clerical
4	13-1151	11-3131	Training and Development Specialists	Training and Development Managers
5	43-4041	13-2041	Credit Authorizers, Checkers, and Clerks	Credit Analysts
6	11-9033	21-1012	Education Administrators, Postsecondary	Educational, Guidance, and Career Counselors and Advisors
7	11-9151	21-1093	Social and Community Service Managers	Social and Human Service Assistants
8	13-1141	11-3111	Compensation, Benefits, and Job Analysis Specialists	Compensation and Benefits Managers
9	11-1011	19-3041	Chief Executives	Sociologists
10	15-2031	19-1042	Operations Research Analysts	Medical Scientists, Except Epidemiologists
11	11-9033	19-3041	Education Administrators, Postsecondary	Sociologists
12	13-1199	11-9199	Business Operations Specialists, All Other	Managers, All Other
13	13-1023	11-3061	Purchasing Agents, Except Wholesale, Retail, and Farm Products	Purchasing Managers
14	49-9043	17-3026	Maintenance Workers, Machinery	Industrial Engineering Technologists and Technicians
15	19-3099	15-2031	Social Scientists and Related Workers, All Other	Operations Research Analysts
16	21-1012	43-4111	Educational, Guidance, and Career Counselors and Advisors	Interviewers, Except Eligibility and Loan
17	11-9161	55-1019	Emergency Management Directors	Military Officer Special and Tactical Operations Leaders, All Other
18	11-1021	43-1011	General and Operations Managers	First-Line Supervisors of Office and Administrative Support Workers
19	49-9012	17-3023	Control and Valve Installers and Repairers, Except Mechanical Door	Electrical and Electronic Engineering Technologists and Technicians
20	11-9033	21-1099	Education Administrators, Postsecondary	Community and Social Service Specialists, All Other
21	11-9161	13-1199	Emergency Management Directors	Business Operations Specialists, All Other
22	19-1042	11-9121	Medical Scientists, Except Epidemiologists	Natural Sciences Managers
23	51-9151	17-3029	Photographic Process Workers and Processing Machine Operators	Engineering Technologists and Technicians, Except Drafters, All Other
24	47-1011	11-9021	First-Line Supervisors of Construction Trades and Extraction Workers	Construction Managers
25	43-9041	13-1031	Insurance Claims and Policy Processing Clerks	Claims Adjusters, Examiners, and Investigators

Source: Lightcast, Authors' calculations
Note: This *Table enumerates the top 25 pairs of SOC6 occupations that exhibit the closest relationship in terms of their skills but differ in the broader SOC2 classification for the year 2014. The low Euclidean distances suggest high similarity, which could indicate overlapping skills or comparable work activities even if they are classified in separate occupational groups.*

For the final validity test we explore whether pairs of occupations that we measure to be closer together tend to have higher bilateral labor mobility. To do that, we regress the computed distance d_{ij} with mobility data from Schubert et al (2024) and from the CPS. Schubert et al (2024) use online worker profiles to estimate labor transitions for pairs of occupations and provide estimates of bilateral transition shares for each of 840 SOC 6-digit origin occupation to each 6 digit destination occupation. These estimates are averages for the period 2002-2016 at the national level, with the majority of observations in the later years. We also confirm these results using data of transition shares for each occupation pair in CPS for the same time period, preceding our sample of distance metrics. The results from CPS are reported in column 5 and because CPS reports occupations at a more aggregated level (305 levels), the number of occupation pairs falls from 696390 to 93025. In CPS, transition shares are calculated using the current occupation reported by each worker and the occupation of the previous year as reported by each worker the day of the survey (longitudinal measurement). The results are reported in Table 3 below. Bilateral distance is negatively correlated with measures of transitional shares across the different sample, and at the more granular level (columns 1-4) this negative correlation with occupational distance is larger, compared to the more aggregated level(column 5). In addition, when pairs of occupations with zero transitions are excluded(column 4), the estimated coefficient becomes even larger.

Table 3 – Regression of bilateral occupational distances on labor mobility

	Dependent Variable: Transition shares between Occ I and Occ j				
	(1)	(2)	(3)	(4)	(5)
Distance	-0.0607***	-0.0533***	-0.0595***	-0.1055***	-0.0324***
	(0.0016)	(0.0015)	(0.0016)	(0.0034)	(0.0036)
Same		0.0033***	0.0036***	0.0053***	0.0025***
		(0.0002)	(0.0003)	(0.0004)	(0.0005)
Distance * Same		-0.0151***	-0.0168***	-0.0266***	-0.0052**
		(0.0012)	(0.0014)	(0.0029)	(0.0023)
Occupation 1 FE	Yes	Yes	Yes	Yes	Yes
Occupation 2 FE	Yes	Yes	Yes	Yes	Yes
Exclude 0s				Yes	
Data		Schubert et al (2024)			CPS
Observations	696,390	696,390	666,672	275,733	93,025
R2	0.32601	0.32931	0.34109	0.37861	0.06697

Robust standard errors clustered by occupation 1 in parenthesis, *** p<0.01, ** p<0.05, * p<0.1

The three validity tests presented above demonstrate that the proposed measure of distance produces robust and reasonable estimates. These results help address valid concerns about the quality and consistency of the OJV data over time.

One potential concern is that the description of tasks in job postings may become more detailed over time, or that certain skills may be replaced by newer equivalents while the underlying tasks remain unchanged (e.g., replacing *TensorFlow* with *PyTorch*). In both scenarios, the computed distances between occupations could artificially increase, making it more difficult to detect convergence.

Another concern relates to the increasing use of standardized templates by HR departments, which may make job descriptions appear more homogeneous. While such a trend could lead to observed convergence in occupational distances, we argue that this reflects a genuine alignment of skill requirements. After all, companies are unlikely to post job descriptions with skills they do not genuinely require, especially if these align with broader industry trends.

We also check robustness of the results by replicating the analysis using *Cosine Similarity* instead of *Euclidean Distance* to assess similarity. The results are identical and omitted for brevity.

Another measure of robustness is to test the effect of using alternative vector sizes when computing bilateral distances. As explained in the **Methodology** section, the skill data provided by Lightcast include approximately 16,000 keywords or phrases. However, since most of these skills are not relevant to most occupations, we represent each occupation at time *t* as a vector of its most important skills. Specifically, we use the top **200 skills** (ranked by importance) for comparison across occupations.

To assess the robustness of this choice, **Figure 6** presents the kernel density functions for bilateral distances in 2014 using alternative vector sizes (e.g., 25, 50, 100, ..., 1,000 skills). The results demonstrate that the density functions remain nearly indistinguishable when using 50 or more observations from the skills vector. This suggests that the choice of vector size does not significantly impact the computed distances. Given these findings, we use **200 observations** to balance computational efficiency with the accuracy of the estimates, as the quality of the results is not compromised.

Figure 6 – Kernel density functions of bilateral distances, 2014, various skill-vector sizes

Finally, to ensure that the results are not driven by outliers, we drop the top 10 or top 25 skills by importance and repeat the analysis. Again, the results remain unchanged.

5.3 Occupations as Networks

A significant advantage of having bilateral distances is their utility as parameters in **gravity-type models** to study mobility patterns. Beyond this, these distances enable the construction of **occupational networks**, where occupations are spatially positioned based on their skill similarities. Using network analysis, we can identify clusters of closely related occupations, pinpoint central roles within the network, and trace pathways for skill transfer. This approach offers deeper insights into the interdependencies among occupational roles and serves as a valuable tool for understanding workforce dynamics and informing policy decisions.

To illustrate this, we construct a **minimum spanning tree (MST)** using the bilateral skill-based distances between SOC4 occupational groups in 2014. The MST minimizes the total edge weight in a graph while ensuring that all nodes are connected. The resulting graph, shown in **Figure 4**, provides a novel visualization of the relationships between occupations.

A closer examination of the network reveals occupations that occupy central positions within the structure. For instance, **Business and Management Specialists** and **Business Operations Specialists** exhibit the highest degree centrality, reflecting their prominence in the occupational landscape. We summarize the top 5 and bottom 5 occupations by degree

centrality in **Table 4** for 2014 and offer a comparison with 2022 data to highlight any changes over time.

To assess the alignment of our network with existing occupational taxonomies, we further **color-code occupations by SOC2 groups**. This allows us to evaluate how well traditional classifications capture skill-based similarities, providing additional validation for our methodology.

Table 4: SOC2 Occupation ranked by degree centrality, 2014 and 2022

Rank	2014	2022
		Top 5
1	Other Management Occupation (SOC2: 11)	Other Management Occupation (SOC2: 11)
2	Business Operations Specialist (SOC2: 13)	Military Enlisted Tactical Operations and Air Weapons Specialists (SOC2: 55)
3	Supervisors of Farming, Fishing, and Forestry Workers (SOC2: 45)	Extraction Worker (SOC2: 47)
4	Health Technologists and Technician (SOC2: 29)	Business Operations Specialist (SOC2: 13)
5	Entertainment Attendants and Related Worker (SOC2: 39)	Top Executive (SOC2: 11)
		Bottom 5
1	Motor Vehicle Operator (SOC2: 53)	Water Transportation Worker (SOC2: 53)
2	Rail Transportation Worker (SOC2: 53)	Other Transportation Worker (SOC2: 53)
3	Military Officer Special and Tactical Operations Leader (SOC2: 55)	Material Moving Worker (SOC2: 53)
4	First-Line Enlisted Military Supervisor (SOC2: 55)	Military Officer Special and Tactical Operations Leader (SOC2: 55)
5	Military Enlisted Tactical Operations and Air Weapons Specialists (SOC2: 55)	First-Line Enlisted Military Supervisor (SOC2: 55)

Figure 7 - Minimum spanning tree of distances among SOC4 occupations

6. Implications of Convergence on Wages and Mobility

So far, we have introduced a novel methodology for measuring skill-based similarities across occupations (or any other groups of interest) and applied it to compute bilateral distances for all pairs of the 1,326 SOC6 occupation definitions. Our findings demonstrate that, over time, occupations are converging as the skill distances between them decrease.

In this section, we empirically demonstrate that occupational convergence has resulted in **wage convergence across occupations** and **higher mobility**. Drawing on the Lucas and Prescott *island model* of occupations, we hypothesize that as islands (occupations) move closer together, the cost of transitioning from one island to another decreases. As a result, we expect to observe two key outcomes: (i) **wages converge** (wage inequality across occupations declines), and (ii) **mobility increases** as workers find it easier to transition between occupations.

Additionally, in **Appendix B**, we examine changes in **within-occupation wage inequality** as occupations undergo transformation. To quantify how much an occupation has transformed

over time, we compute its **self-distance**—the bilateral distance of the occupation relative to itself—at a benchmark year (e.g. 2014) and a subsequent year (e.g. 2022). Our findings suggest that occupations experiencing **greater transformation** (i.e., higher self-distance between 2014 and 2022) also exhibit higher wage growth and higher reductions in *within-occupation* wage inequality.

6.1 Wage convergence

We begin by investigating whether the convergence of occupations over time has led to convergence in wages. To explore this relationship, we plot in **Figure 8a** the average computed bilateral distances between occupational pairs at the SOC4 level alongside the **coefficient of variation** of the median wage over time. The results clearly illustrate that since 2014, as occupational distances decrease—indicating greater skill convergence—wages have also converged. This observed relationship supports the hypothesis that as occupations become more similar, wage disparities across occupations decline.

Figure 8 - CV and average distance over time

(a) Average occupational distance (SOC4) and wage inequality (coefficient of variation)

(b) Average occupational distance (SOC4) and labor mobility

Source: Lightcast, Authors' calculations

Note: This figure in panel a. plots the time-series trends of wage variation and occupational dissimilarity from 2014 to 2022. The declining coefficient of variation suggests a convergence of wages within SOC6 occupation groups, while the decline in median Euclidean distance trend suggests an increase in occupation homogeneity. The similar trends observed in recent years between wage inequality and occupational distance is not present when wage variation and distance are calculated by MSAs (in panel b).

To formalize this finding and address concerns about omitted variable bias, we conduct a regression analysis. Specifically, we regress the change in relative wages between occupations *i* and *j* on the change in bilateral distances d_{ij} for the years 2014 and 2022. This approach strengthens our test of the hypothesis by moving beyond aggregate trends over time. Instead, it leverages the richness of the bilateral data to determine whether wage convergence is greater for occupations that become more similar in terms of skills.

We compute **relative wages** between occupations *i* and *j* as follows:

$$\Delta rw_{ij} = \left[rw_{ij}^{2022} - w_{ij}^{2014}\right] = \left|\left(\frac{w_i^{2022} - w_j^{2022}}{w_i^{2022} + w_j^{2022}}\right) - \left(\frac{w_i^{2014} - w_j^{2014}}{w_i^{2014} + w_j^{2014}}\right)\right|$$

The regression results are presented in **Table 5**. In Column 1, we control for the initial wage gap in 2014 (rw_{2014}) and the change in bilateral distance (Δeuc) between occupations from 2014 to 2022. Here, a higher value of Δeuc indicates that the two occupations have moved further apart. The results reveal two key findings: first, the larger the initial wage gap in 2014, the greater the wage convergence observed. Second, and most importantly, occupations that have become closer in terms of skill similarity exhibit significantly higher wage convergence.

In Column 2, we extend the analysis by controlling for **self-distance**, which measures how much an occupation *i* has transformed between 2014 and 2022. The results remain robust: occupations that converge in skills also experience greater wage convergence. Additionally, we find that occupations undergoing higher self-transformation also exhibit greater wage convergence, suggesting that internal transformation plays a role in reducing wage disparities.

In Columns 3 and 4, we extend the analysis by introducing further splits in the data. In Column 3, we differentiate between occupation pairs *i* and *j* that belong to the same SOC2 group and those that do not. In Column 4, we explore whether occupation *i* is classified as *Low Non-Routine Cognitive* (Low NR-C) or *High Non-Routine Cognitive* (High NR-C) and interact this classification with our distance variable.

The results reveal that, over the period 2014–2022, occupations classified as *Low NR-C* experienced the largest reduction in wage inequality. This finding contrasts with the pre-2010 trend of increasing inequality for these occupations, as documented in previous studies. For *Low NR-C* occupations, the initial level of wage inequality has a smaller effect on the observed dynamics of convergence. However, the effect of skill-based distance becomes more pronounced in magnitude for occupations outside the non-routine cognitive classification.

These results highlight the pivotal role of skill convergence in driving wage convergence, particularly for occupations that are not classified as non-routine cognitive.

6.2 Labor mobility

We also investigate the impact of occupational convergence on labor mobility. To begin, we plot the average skill-based distance between occupations alongside the average labor mobility from 2014 to 2022 in Figure 8b. The results show a clear pattern: as occupational

Table 5 - Occupations convergence and wage convergence regressions, 2014 and 2022

	(1)	(2)	(3)	(4)
rw2014	-0.101***	-0.101***	-0.102***	-0.109***
	-0.000489	-0.000489	-0.000503	-0.000789
Same			-0.00479***	
			-0.000563	
Same * rw2014			0.0207***	
			-0.00268	
Δdist	0.0819***	0.0781***	0.0813***	0.0112***
	-0.00123	-0.00126	-0.00126	-0.0025
Same*Δdist			0.00872	
			-0.00558	
Self_distance		-0.0190***		
		-0.00134		
Low NR-C				-0.00830***
				-0.000327
Low NR-C * rw2014				0.00652***
				-0.00106
Low NR-C * Δdist				0.0774***
				-0.0032
Constant	0.0100***	0.0113***	0.0104***	0.0143***
	-0.000148	-0.000174	-0.000155	-0.000249
Observations	497,730	497,730	497,730	419,475
R-squared	0.088	0.088	0.088	0.092

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

Source: Lightcast, Authors' calculations

Note: This table presents the estimated coefficients from a regression examining the relationship of the dependent variable (change in wage inequality) with initial wage inequality (2014), occupation dissimilarity, occupational characteristics, and their interactions with occupation dissimilarity across four different model specifications. The models consistently demonstrate significant effects for initial wage inequality (rw2014) and the change in distance (Δdist) on inequality. Additionally, interaction terms reveal differential impacts when considering occupational similarity and nonroutine cognitive task intensity.

distances decrease—indicating greater skill convergence—labor mobility rises. This suggests that workers find it easier to transition between occupations as the skill requirements become more similar over time.

To formally test the relationship between occupational convergence and labor mobility, we regress the **change in distance** (Δd_{ij}) between two occupations *i* and *j* on the **change in transitional shares** between the same occupation pairs for the years 2014 and 2022. The results are presented in **Table 6**. In *Column 1*, we include fixed effects for occupation *i*, occupation *j*, and their interaction *ij*. The coefficient of Δd_{ij} on labor mobility is negative and statistically significant, indicating that occupations moving closer in terms of skill-based distance experience larger increases in labor mobility. In *Column 2*, we further control for

changes in self-distance—a measure of how much each occupation has transformed over time. In *Column 3*, we introduce employment-based weights to account for the relative size of each occupation. Importantly, the results remain robust across all specifications: pairs of occupations that converge (i.e., exhibit a reduction in d_{ij}) experience a significant increase in labor transitions.

Table 6 - Occupations convergence and labor mobility regressions, 2014 and 2022

	Dependent Variable: Change in Transition shares between Occ I and Occ j		
	(1)	(2)	(3)
Change in Distance	-0.1504***	-0.1213**	-0.1172**
	(0.0575)	(0.0579)	(0.0556)
Same		-0.0127	-0.0111
		(0.0137)	(0.0132)
Change in Distance* Same		-0.0464	-0.0430
		(0.0338)	(0.0330)
Occupation 1*Year FE	Yes	Yes	Yes
Occupation 2*Year FE	Yes	Yes	Yes
Occupation 1*Occupation 2 FE	Yes	Yes	Yes
Weighted by log employment	No	No	Yes
Observations	24,159	24,159	24,159
R2	0.78044	0.78052	0.77131

Source CPS. Robust standard errors clustered by occupation 1 in parenthesis, *** p<0.01, ** p<0.05, * p<0.1

7. Conclusion

In this study, we introduced a novel methodology for measuring skill-based distances between occupations and demonstrated its wide-ranging applications. By leveraging detailed skill data from job advertisements, we computed bilateral distances between 1,326 SOC6 occupations and observed a clear trend of convergence over time (2014–2022), where occupations have become more similar in their skill requirements.

We empirically showed that this occupational convergence has significant implications for labor market outcomes. First, wage convergence accompanies occupational convergence. Initial gaps in wages narrow as skill distances decrease, and this relationship holds even after controlling for initial wage disparities, self-distances, and other factors. Notably, occupations classified as Low Non-Routine Cognitive (Low NR-C) experienced the largest reduction in wage inequality, reversing earlier trends of rising inequality in these roles.

Second, we demonstrated that occupational convergence leads to higher labor mobility. As skill-based distances between occupations decrease, workers find it easier to transition between roles, increasing overall mobility within the labor market. Regression analysis confirmed this relationship: the reduction in bilateral distances is significantly associated with rising transitional shares between occupations, even after accounting for fixed effects and employment-based weights.

Beyond these core findings, our methodology enabled us to construct occupational networks using skill-based distances. By employing tools such as minimum spanning trees and centrality measures, we visualized the interdependencies among occupations and identified key nodes, such as Business and Management Specialists, which occupy central roles within the occupational landscape. This network-based approach provided further validation of our results and highlighted the practical utility of our distance measures for understanding workforce dynamics.

In summary, the convergence of occupations over time has led to wage convergence and increased labor mobility, signaling a more integrated and dynamic labor market. These findings provide valuable insights into the changing structure of occupations and underscore the importance of skill-based analysis for informing policy decisions related to workforce development, reskilling initiatives, and wage inequality.

Bibliography

1. Acemoglu, Daron. "Directed technical change." *The review of economic studies* 69.4 (2002): 781-809.

2. Acemoglu, Daron, et al. "Artificial intelligence and jobs: Evidence from online vacancies." *Journal of Labor Economics* 40.S1 (2022): S293-S340.

3. Aeppli, Clem, and Nathan Wilmers. "Rapid wage growth at the bottom has offset rising US inequality." *Proceedings of the National Academy of Sciences* 119.42 (2022): e2204305119.

4. Autor, David H., Lawrence F. Katz, and Alan B. Krueger. "Computing inequality: have computers changed the labor market?." *The Quarterly journal of economics* 113.4 (1998): 1169-1213.

5. Autor, David H., Frank Levy, and Richard J. Murnane. "The skill content of recent technological change: An empirical exploration." *The Quarterly journal of economics* 118.4 (2003): 1279-1333.

6. Autor, David H. "Why are there still so many jobs? The history and future of workplace automation." *Journal of economic perspectives* 29.3 (2015): 3-30.

7. Autor, David H., and David Dorn. "The growth of low-skill service jobs and the polarization of the US labor market." *American economic review* 103.5 (2013): 1553-1597.

8. Song, Jae, et al. "Firming up inequality." *The Quarterly journal of economics* 134.1 (2019): 1-50.

9. Cammeraat, Emile, and Mariagrazia Squicciarini. "Burning Glass Technologies' data use in policy-relevant analysis: An occupation-level assessment." (2021).

10. Card, David, and John E. DiNardo. "Skill-biased technological change and rising wage inequality: Some problems and puzzles." *Journal of labor economics* 20.4 (2002): 733-783.

11. Carrillo-Tudela, Carlos, and Ludo Visschers. "Unemployment and endogenous reallocation over the business cycle." *Econometrica* 91.3 (2023): 1119-1153..

12. Chodorow-Reich, Gabriel, and Johannes Wieland. "Secular labor reallocation and business cycles." *Journal of Political Economy* 128.6 (2020): 2245-2287.

13. Deming, David, and Lisa B. Kahn. "Skill requirements across firms and labor markets: Evidence from job postings for professionals." *Journal of Labor Economics* 36.S1 (2018): S337-S369.

14. Kambourov, Gueorgui, and Iourii Manovskii. "Occupational mobility and wage inequality." *The Review of Economic Studies* 76.2 (2009): 731-759.

15. Katz, Lawrence F., and Kevin M. Murphy. "Changes in relative wages, 1963–1987: supply and demand factors." *The quarterly journal of economics* 107.1 (1992): 35-78.

16. Kuhn, Peter, and Mikal Skuterud. "Internet job search and unemployment durations." *American Economic Review* 94.1 (2004): 218-232.

17. Lilien, David M. "Sectoral shifts and cyclical unemployment." *Journal of political economy* 90.4 (1982): 777-793.

18. Lucas Jr, Robert, and Edward C. Prescott. "Equilibrium search and unemployment." *Journal of Economic Theory* 7.2 (1974): 188-209.

19. Modestino, Alicia Sasser, Daniel Shoag, and Joshua Ballance. "Upskilling: Do employers demand greater skill when workers are plentiful?." *Review of Economics and Statistics* 102.4 (2020): 793-805.

20. Moscarini, Giuseppe, and Kaj Thomsson. "Occupational and Job Mobility in the US." *The Scandinavian Journal of Economics* 109.4 (2007): 807-836.

21. Western, Bruce, and Jake Rosenfeld. "Unions, norms, and the rise in US wage inequality." *American Sociological Review* 76.4 (2011): 513-537.

Appendix A: Artificial Intelligence and Data Analytics Skills List

ARTIFICIAL INTELLGIENCE (AI) SKILLS	DATA ANALYTICS (DA) SKILLS
Apache Ant	Apache Hadoop
Apache Spark	Apache Hive
Artificial Intelligence	Apache Kafka
Automated Testing	Bayesian Inference
Automation	Bayesian Methods
Automation Consulting	Bayesian Modeling
Automation Systems	Bayesian Networks
Automation Techniques	Big Data
Automation Tools	Big Data Analytics
AWS Elastic MapReduce (EMR)	Data Analysis
BigQuery	Data Analytics
Boosting (Machine Learning)	Data Conversion
Caffe Deep Learning Framework	Data Engineering
Cluster Analysis	Data Management
Clustering	Data Modeling
Clustering Algorithms	Data Science
Computer Vision	Data Visualization
Convolutional Neural Network (CNN)	Extraction Transformation and Loading (ETL)
Decision Trees	Microsoft Power BI
Deep Learning	Microsoft SQL
Kubernetes	MongoDB
Machine Learning	MySQL
Machine Vision	NoSQL
MapReduce	Oracle PL/SQL
ML/AI	PostgreSQL
Natural Language Processing	Predictive Analytics
Natural Language Toolkit (NLTK)	Predictive Models
Neural Networks	Prepare Spreadsheets
Splunk	Spreadsheets
Supervised Learning (Machine Learning)	SQL
Support Vector Machines (SVM)	SQL Injection
TensorFlow	SQL Plus
Torch (Machine Learning)	SQL Server
	SQL Server Analysis Services (SSAS)
	SQL Server Reporting Services (SSRS)
	SQL*Loader
	SQLAlchemy
	SQLite
	Tableau
	Transact-SQL

Appendix B: Self-Distance and Within-Occupation Changes

B1 Measuring occupational transformation (self-distance)

Our methodology allows us to measure how each occupation has **transformed over time**. To quantify this, we compare the skills-vector representation of an occupation *i* at a benchmark year *t* with its representation at a later year *t+T*. The degree of transformation is captured by the distance (or dissimilarity) between these two vintages of the same occupation. A higher distance indicates a greater transformation in the skills associated with that occupation.

We refer to this measure as **self-distance**, which reflects the extent to which an occupation's skill requirements have evolved over time. This metric provides a straightforward yet powerful way to track occupational transformation and assess its implications for wage dynamics, inequality, and mobility.

For each occupation, we calculate the self-distance annually, using 2014 as the benchmark year. In **Figure B1**, we report the median self-distance across all occupations for each year. To provide further insights, we replicate this analysis by SOC6 type, distinguishing between *High Non-Routine Cognitive* (High NR-C) and *Low Non-Routine Cognitive* (Low NR-C) occupations.In the figure, the middle line represents the median self-distance for all SOC6 occupations combined, while the top line corresponds to Low NR-C occupations and and the **bottom line** represents High NR-C occupations. The results show that occupations are transforming over time, with **Low NR-C occupations changing more rapidly** compared to High NR-C occupations.

Figure B1 – Median Euclidean Self-Distance by Year and Occupation Type (High vs Low Nonroutine Cognitive)

Source: Lightcast, Authors' calculations

Note: This figure compares the median self-distance within occupation types characterized by high and low levels of nonroutine cognitive tasks across different years. The self-distance measure reflects the average within-occupation homogeneity of occupations, with a lower median self-distance indicating a higher similarity.

B2 Self-distances, wage inequality, and wage growth

We also examine the potential relationship between the degree of occupational transformation from 2014 to 2022 and changes in within-occupation wage inequality and wage growth.

To measure within-occupation wage inequality, we compute two metrics for each SOC6 occupation group in each year:

- The **p75/p25 ratio**, defined as (p75−p25)/p50(p75 - p25) / p50(p75−p25)/p50, where *p* represents the percentile of the wage distribution.
- The **p90/p10 ratio**, which compares wages at the 90th and 10th percentiles.

In **Figure B2**, we plot the average within-occupation inequality across all SOC6 groups for each year, using the *p75/p25* ratio on the left and the *p90/p10* ratio on the right.

The results reveal that regardless of the inequality measure used, within-occupation wage inequality has decreased over time. This decline coincides with the observed transformation of occupations, suggesting a potential connection between the two trends.

Figure B2 – Wage Inequality Within SOC6 Occupations

Next, we regress the change in within-occupation wage inequality (p75/p25) between 2014 and 2022 on the self-distance of each occupation, with results presented in **Table B1**. In Column 1, we analyze all SOC6 occupations, while in Column 2, we further decompose occupations into High NR-C and Low NR-C groups.

The regression results provide important insights into the dynamics of wage inequality within SOC6 occupational groups over the study period. The significant negative coefficients for **Ineq_2014** in both columns (-0.190 and -0.192) indicate that higher initial levels of inequality in 2014 were strongly associated with larger reductions in within-occupation inequality by 2022. This result, statistically significant at the 1% level, suggests a convergence trend within occupations, supporting our earlier findings on decreasing wage inequality.

Table B1 – Within SOC6 wage inequality and occupational transformation

	(1)	(2)
Ineq_2014	-0.190***	-0.192***
	-0.0241	-0.0376
Low NR-C		-0.042
		-0.0301
Low NR-C * Ineq_2014		-0.0687
		-0.0505
Self Distance	-0.147**	-0.333**
	-0.0605	-0.132
Low NR-C * Self Distance		0.290*
		-0.161
Constant	0.0578***	0.102***
	-0.0143	-0.0231
Observations	753	637
R-squared	0.08	0.17

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

Source: Lightcast, Authors' calculations

Note: This table presents the estimated coefficients from a regression examining the relationship of initial wage inequality, occupational dissimilarity and within-occupation wage variability. The models consistently demonstrate significant effects for initial wage inequality and the change in distance (Δdist) on the variability of wages within an occupation.

The variable **Self Distance**, which measures the degree of occupational transformation, also shows significant negative coefficients in both columns. This implies that occupations undergoing greater task transformations over time tend to experience larger reductions in within-occupation inequality.

In Column 2, the interaction term **Low NR-C * Self Distance** is positive (0.290) but only significant at the 10% level. This suggests that the relationship between self-distance and wage inequality reduction may differ for Low NR-C occupations. Interestingly, this indicates that **non-routine cognitive occupations** tend to exhibit the most substantial changes in within-occupation wage inequality. However, the overall results suggest that the convergence trend within occupations is not heavily dependent on the nature of tasks, whether they are routine or cognitive.

We also regress self-distance on wage growth and present the results in **Table B2**. We find that occupations that had lower median wages (**Med wage 2014**) in 2014 experienced lower growth. However, **LOW NR-C** and occupations with higher **Self Distance** measure experienced higher wage growth between 2014 and 2022.

Table B3 – Within SOC6 wage growth and occupational transformation

	(1)	(2)
Med wage 2014	-1.82e-06***	-5.42e-07**
	(1.58e-07)	(2.44e-07)
Low NR-C		0.251***
		(0.0277)
Low NR-C * Med wage 2014		-4.99e-06***
		(5.11e-07)
Self Distance	0.247***	0.367***
	(0.0624)	(0.122)
Low NR-C *Self Distance		-0.219
		(0.151)
Constant	0.290***	0.188***
	(0.0108)	(0.0198)
Observations	651	547
R-squared	0.222	0.366

Standard errors in parenthesis, *** p<0.01, ** p<0.05, * p<0.1

These findings highlight the critical role of occupational transformation in shaping wage dynamics and suggest that task evolution is a key driver of reduced wage inequality within occupations.