Artificial Intelligence's Effects on Skill Premium Changes in the United States

By Malisha Weerasinghe, Swarna Wijaya, and Theo Obadiah Teguh

Artificial intelligence (hereafter AI) is on the mandate of almost every organization today, transforming productivity and redefining the boundaries of work (Singla et al., 2025). Almost 60% of jobs in advanced economies are expected to be impacted by AI (IMF, 2024) - yet the sentiment towards how workers and their tasks will be affected remains ambiguous. This paper aims to assess how AI shapes the skill premium through analysis of changes in wage inequality as measured by the Theil Index. Our regression results reveal that the scale and complexity of AI models (measured by AI model parameter counts) exacerbate wage gaps, especially among low-skilled workers, but AI research and development (using AI patent filings for proxy) show no significant short-term effect on inequality. This suggests that technological advances alone do not immediately translate into labor market disparities, but the complexity of models and the problems they are able to solve may disproportionately disadvantage lower-skilled workers. Following an analysis of conservative, baseline, and optimistic projections of AI parameter growth over 5 years, we anticipate that wage inequality is likely to continue harming low-skilled workers more than their high-skilled counterparts even in the future. Employment growth also has significant implications on disparities between occupations, likely due to skill-biased demand. *Keywords: AI, Wage Inequality, Skill Premium, United States, Occupations, Bureau of Labor Statistics*

Artificial intelligence (hereafter AI) is on the mandate of almost every organization today, transforming productivity and redefining the boundaries of work (Singla et al., 2025). Almost 60% of jobs in advanced economies are expected to be impacted by AI (IMF, 2024) - yet the sentiment towards *how* workers and their tasks will be affected remains ambiguous. Although some experts anticipate that AI will completely take over jobs, others assert that it will be a complement to workers more than a substitute, increasing productivity (Rosalsky, 2025). In a market economy where earnings are largely tied to productivity and skill (Autor, 2022), the uncertainty of the impact of AI on these factors translates to an uncertainty of income distribution among workers. It is this possible distributional heterogeneity that we aim to examine.

Our study focuses on the United States, justified by three factors as follows.

Firstly, the US has been at the forefront of AI development for several years. The US had started rigorous research on AI more than a decade before China did (Lee, 2018). Between 2013 and 2023, American AI-related grants increased by almost 40% annually (Stanford HAI, n.d.). As a result, 47% of US jobs may get replaced by AI within the next decade (Frey and Osborne (2023), as cited in Lee (2018)). Such evidence proves the relevance and direct impact of AI on the job market in the US. With the abundance of data and empirical results, the US is an ideal country for this study.

Secondly, given the direct influence of AI in the workplace, the US has enacted several relevant policies to minimize its negative spillovers. In 2024, the US Department of Labor (DOL) introduced an AI Best Practices Roadmap to ensure that the adoption of AI is well maximized and inclusive for all workers affected by AI use (Department of Labor, n.d.*b*). They realized that the opportunities presented by AI were not equally reaped by all workers. This guideline was then expected to uplift workers' rights, employment stability, and workers standard. The release of these approaches has shown growing concerns from policymakers on controlling the effect of AI. Our paper then hopes to be an aid to policymakers, be it in the US or anywhere else in the world, to reinforce the importance of formulating policies relevant to development of AI.

Thirdly, wage inequality analysis is heavily dependent on the availability of the country's labor data. In this regard, the US provides publicly accessible data on wages and employment. Specifically, the US Bureau of Labor Statistics (BLS) releases annual datasets on the average income and employment level of job occupations. This allows us to explore the effect of AI on specific wage inequality based on their skill levels, according to the International Labour Organization (ILO)'s standards.

In terms of time range, we utilize data from 2020 to 2024 for our main regression analysis. Our research looks at the percentage change of the variables for which we retrieve data from 2019 to 2024. For this particular scope, our rationale lies on two main factors.

Firstly, the most relevant AI only became more evident during this period. This is accelerated by the strong urgency from the COVID-19 pandemic for corporations to implement digital solutions. Based on a survey conducted by Balakrishnan et al. (2020), one-fifth of its respondents reported that AI contributes to at least 5% of their bottom-line.

Secondly, current labor wage datasets are reported using the 2018 Standard Occupation Classification (SOC) from the Bureau of Labor Statistics, valid for data between 2018 and 2024 (Bureau of Labor Statistics, 2025). For simplicity, we chose to maintain this standard and avoid using data from previous years, which follow older classifications. Therefore, we focus on collecting data from 2019 and thoroughly analyzing them in the 2020-2024 scope.

Given this context, this report aims to assess how AI development in the US shapes skill premium through analysis of changes in wage inequality, measured through a modified Theil index, over time.

We utilize the Theil index primarily for its subgroup decomposability. The index allows us to isolate how much inequality stems from disparities between skill levels versus within them, in turn providing more light to AI's skill-biased effects (Georgieff, 2024). This separability allows us to single out each group's contribution to total inequality as well. The calculation also ensures the robustness of our model, mitigating the ambiguity of the conventional Gini coefficient (World Bank, n.d.).

We proxy AI development via two distinct channels: AI model parameters and AI patents. In simple terms, AI parameters account for the quality of AI performance (TEDAI San Francisco, n.d.), while AI patents account for the quantity of AI research and development. The performance of AI models depends on scale, determined by the number of model parameters, the computational cost of training, and the size of the dataset (Kaplan et al., 2020). For example, models like GPT-4 are estimated to include 1.8 trillion parameters, trained on 13 trillion tokens, with a compute cost of $65 million (Li, 2020; Brown et al., 2020; Patel and Wong, 2023, as cited in Brynjolfsson, Li and Raymond (2025)). For the purpose of our research, we isolate AI parameters as a measure of the scale of the AI model, effectively acting as a translation of the computing capabilities of the model, which is likely to enhance high-skill work (Our World in Data, 2025). This metric was used partially due to the availability of the largest public dataset of its kind through Epoch AI (2025), and the 'easily estimable numbers from descriptions of model architecture' (Sevilla, Villalobos and Cerón, 2021). The second proxy of AI development is AI patent filings, reflecting R&D activity and proprietary innovation. Maher and Schaffelke (2023) state that patents are widely considered early indicators of innovation, and the main forces of economic growth and development. Given that businesses are increasingly investing more into AI development (Brynjolfsson, Li and Raymond, 2025), it was a clear metric to use in our analysis.

Utilizing data from the BLS and associated crosswalks from the ILO (2025), we categorize occupations into their skill level - low, middle, and high-skill - and assess their associated changes in income. In addition, we also utilize feature engineering to handle ambiguous or missing data. Other control variables, including Gross Domestic Product (GDP) and Personal Consumption Expenditures (PCE), as well as employment, are utilized to prevent omitted variable bias. These variables are then added to our fixed effects regression model, which isolates the effect of AI while controlling for skill levels, macroeconomic trends, and occupation-specific heterogeneity.

Our regression results show that the deployment of AI through model parameters exacerbates wage gaps. Holding other variables constant, a 100,000 percentage point increase in AI parameters leads to a 0.12 percentage point increase in the Theil Index on average, significant at the 1% level. The complexity of AI models is likely an aid in higher-level thinking tasks, most commonly done by high-skilled workers. In contrast, we find that AI patents have little effect on inequality, likely because of the delayed effect of AI patents being deployed by businesses. It is possible that the patents themselves take time to have an effect on businesses as a result. Further results from our regression also suggest that employment growth has significant implications for the disparities between occupations. This is likely due to skill-biased demand, as the skill premium increases demand for high-skilled workers. Changes in employment levels had a positive and statistically significant impact on wage inequality - an increase in employment changes by 100,000 percentage point is associated with an increase of the Theil Index by 183.724 percentage point on average.

As an extension of our analysis, we extrapolate the performance of AI (through predicted increases of AI parameter count) to provide 1, 2, and 5-year projections of AI's impact on wage inequality within skill levels, and then between skill levels. We utilize linear, quadratic, and exponential growth potentials of AI parameter counts to correspond to conservative, baseline, and optimistic projections respectively. We find that the conservative growth of AI complexity will lead to 3.42% less inequality growth within high-skilled occupations compared to low-skilled occupations in the next five years. In the optimistic expectation that AI models become more complex and increase in the number of parameters, wage inequality growth will increase by almost three times in the span of five years.

Our results are reflective of previous literature surrounding this topic, which has mostly centered on the implications of computerization and automation via robots, or utilizes specific language models in individual case studies. Theoretically, we identify two main frameworks to explain wage disparities caused by technology - Skill-biased Technological Change (SBTC), and the task polarization model. SBTC explains that wage disparities are a consequence of high-skilled workers being able to benefit substantially from technological change compared to their low-skilled counterparts. On the other hand, the task polarization model singles out specific low-skill, routine tasks that can be automated and consequently replace low-skilled workers. Regardless of the framework used, both theories imply that technological change creates an inevitable wage disparity between low- and high-skilled labor.

In comparison, some case-specific research shows mixed results. A study by Brynjolfsson, Li and Raymond (2025) showed that low-skilled workers were able to learn from AI's recommendations, where the model is trained in best practices, creating a 14% increase in productivity on average. Other research showed no treatment effect of access to AI on average (Otis et al., 2024).

Other research on AI and its effect on skill premium has also been done by Bloom et al. (2024), who explore wage disparity through a production function, showing that AI reduces skill premium.

However, this research paper, to our knowledge, is the first of its kind to analyze the implications of AI based on income data and the Theil Index on a macroeconomic scale. We make use of specific data sets from the Bureau of Labor Statistics in pursuit of visible results within the economy.

Our paper is organized as follows. Section 1 will outline our data handling methodology alongside our proposed regression model. Following the testing of our hypothesis, Section 2 will discuss our results from our main regression, outlining any variables that were statistically significant in the discussion of AI on changes in wage inequality. We extend this discussion with results from extrapolation of AI parameter data, allowing us to project future implications on wage inequality. This will be followed by corroborating evidence and relevant literature in Section 3. Finally, in Section 4, we discuss extensions and possible topics for future research, followed by our concluding remarks in Section 5.

I. Data Handling and Methodology

A. Data Sources

I.A.1 Income per Occupation

The US Bureau of Labor Statistics (hereafter, BLS) lists out employment numbers, as well as average and median hourly wage, for approximately 850 occupations each year (Bureau of Labor Statistics, 2025). These occupations are categorized based on the 2018 Standard Occupational Classification (SOC) system, a federal statistical standard by the BLS. Each of the 850 or so occupations is classified as 'detailed' and each detailed occupation is placed in one of 459 broad occupations. The occupations are then further subdivided to 98 minor groups, and 23 major groups to facilitate classification (Bureau of Labor Statistics, 2025). These groupings are based on similar job duties, skills, education, and/or training per detailed occupation.

The number of occupations provided each year had some variation, as some occupations were added or removed by the BLS. To ensure consistency in the time panel data, we considered 2024 detailed occupations as the baseline for our research. For detailed occupations that did not exist in the 2024 detailed occupation list, associated occupations from previous years were removed. Those included: 'Dancers', 'Actors', 'Musicians and Singers', 'Disc Jockeys, Except Radio', 'Entertainers and Performers, Sports and Related Workers, All Other'.

For missing data entries in income, the Little's MCAR test was first conducted to justify the use of an imputation method and a Multivariate Normal test to ensure the validity of normal distributional assumptions. Failure to reject the null hypothesis of the MCAR test justified the use of the Multiple Imputation by Chained Equations - Predictive Mean Matching (MICE-PMM) algorithm to impute any missing values to create a complete dataset (see A.A1 for detailed MICE imputation process). In simple terms, the MICE-PMM algorithm allowed us to predict missing income in a detailed occupation utilizing data from previous years. With an imputed and completed dataset, the annual percentage changes of income per occupation were calculated.

I.A.2 Employment per Occupation

Employment numbers per occupation are also provided by the BLS (2025) for the years 2019 to 2024, and were used as a control factor to account for any labor demand shifts. This allowed us to isolate AI's direct effect on wages as opposed to its effects on job creation or replacement.

The data were handled in a similar way to income data (Section 2.1.1), where we used 2024 as the baseline year and the MICE-PMM algorithm to impute missing values. We calculated annual percentage changes to employment numbers per occupation and combined these data with their associated income changes.

I.A.3 Skill Level Classification

SOC 2010 and ISCO-08 Crosswalk

Each detailed occupation provided by the US BLS is classified by a Standard Occupational Classification System (SOC or OCC Code), which is renewed periodically. The detailed occupations we utilized were classified by the SOC 2018 standard. For skill classifications, SOC occupations can be mapped to the International Labor Organisation's Standard Classification of Occupations (ISCO-08 Codes) (2024), which in turn have Skill Classifications from Level 1 (Low) to Level 4 (Very High).

We found that crosswalks of ISCO-08 and 2010 SOC codes are provided directly by the BLS, but crosswalks to ISCO-08 and 2018 SOC codes were unavailable. A sample of the crosswalk is provided in Table 1 below:

ISCO-08 Code	ISCO-08 Title	Skill Level (1-4)	2010 SOC Code	2010 SOC Title
1111	Legislators	3 and 4	11-1031	Legislators
1112	Senior government officials	3 and 4	11-1011	Chief Executives
1112	Senior government officials	3 and 4	11-1021	General and Operations Managers
1112	Senior government officials	3 and 4	11-9161	Emergency Management Directors

Table 1—ISCO-08 and 2010 SOC Crosswalk

To match the 2018 SOC codes with ISCO-08 Occupation Classifications and their associated skill levels, we used 2010 SOC crosswalks between the two datasets, as shown in Figure 1. In simple terms, we mapped ISCO-08 skill levels to 2010 SOC codes and then 2010 SOC to 2018 SOC.

Figure 1. Visualization of mapping ISCO-08 Skill level to SOC 2018 Classification

Skill level re-classification

The ISCO-08 Skill Classifications range from 1 to 4 (see A1). To simplify our analysis, we merged Level 3 and 4 classifications to one single 'High' skill classification, creating 3 major classifications - Low (Level 1), Medium (Level 2), and High (Level 3&4).

Handling corner cases with RoBERTa

Handling corner cases not directly concluded in the previous indirect crosswalk mapping proved challenging, however, as there were instances of many-to-many mapping. This resulted in multiple rows for an occupation title, but with different skill levels. We extracted a table out of the crosswalk datasets as an illustration below:

2010 SOC Title	**ISCO-08 Title**	**ISCO Skill Level Classification**
Gaming Surveillance Officers and Gaming Investigators	Legal and Related Associate Professionals	High (2)
	Security Guards	Medium (1)

Table 2—SOC Title and ISCO Classification

We used an SVM classifier on RoBERTa word embeddings, imputed with ADASYN, as a tool to perform a tiebreaker for skill level categorization.

These embeddings were trained on the ISCO-08 skill categorization data set, with a standard 80-20 train-test split. We then evaluated the performance of SVM with standard classification metrics and checked for linear separability of the predictions with PCA (see A.A3).

Skill levels are categorized in an ordinal scale of 0 (low-skilled), 1 (medium-skilled), and 2 (high-skilled). We gathered all 44 occupations with multiple skill level classifications post-merge (see A3 & A4). Afterwards, a tiebreaker function handled the following three cases.

1) If the SVM classification is within the range of skill-levels determined by the crosswalk, then the final tiebreaker result is indeed the SVM classification.

2) If the SVM classification is greater than the maximum of the range of skill-levels determined by the crosswalk, then the final tiebreaker result is the maximum of the set of skill-levels.

3) If the SVM classification is less than the minimum of the range of skill-levels determined by the crosswalk, then the final tiebreaker result is the minimum of the set of skill-levels.

When we run this through the above title, for example, 'Gambling and Surveillance Officers and Gambling Investigators' (2018 SOC Title) is classified as Low Skill (0). According to the criteria set above, this job title is then classified as low-skilled in our model.

In addition, there were five occupations that were completely lost during the income-employment and ISCO-OCC-2018 merge (see A5). For these five entries, we directly used the SVM classifier's predicted skill levels.

The aforementioned process allowed us to create a one-to-one link from each available occupation to a single skill level. Our final data set is our detailed occupation data, associated skill levels, along with percentage change of employment and income per occupation.

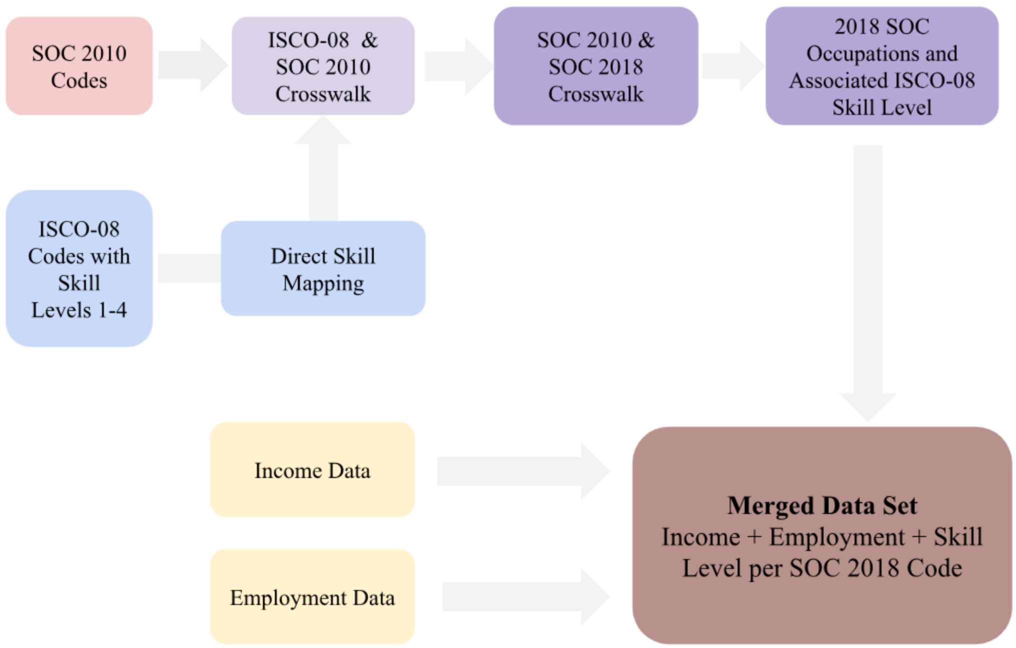

FIGURE 2. VISUALIZATION OF MAPPING OCC CODES TO ISCO CODES FOR SKILL LEVEL ASSIGNMENT

I.A.4 GDP & PCE GROWTH

Annual Gross Domestic Product (GDP) and Personal Consumption Expenditure (PCE) growth are sourced from the Federal Reserve Bank of St. Louis (2025), and are used to control for any economic cycle implications on wage inequality, which is likely owing to the overlap between the time period under consideration and US economic recessions during the COVID-19 pandemic. Both variables are lagged by a year to address temporal dynamics and improve causal inference.

I.A.5 AI PARAMETERS

AI parameter data are included as a measure of AI development and capabilities of AI systems, which in turn may shape how jobs are affected in the US. AI parameters represent developments in model size, computational power, and the complexity of training methodologies (Epoch AI, 2024). These parameters act as a proxy of how capable an AI model is in handling a wider range of complex tasks - a higher parameter count of an AI model means that the number of occupations and tasks that can be impacted by AI increases (Our World in Data, 2025). These data are frequently updated, depending on the development of AI models, by Epoch AI.

For the purpose of our regression, we calculated the average number of parameters (which can range from million to trillion units) produced each year. If there are missing values of AI parameter for certain AI models, they were excluded out of our analysis. We chose to take an average to avoid overexposure to outliers, providing us with the central tendency of AI adoption intensity across the labor market. This would effectively allow us to capture the 'typical' worker's exposure to AI while ensuring comparability across years. To capture marginal productivity effects, we take the annual percentage change of parameter models for the periods under consideration.

I.A.6 AI Patents

AI Patent data are used as a proxy of AI innovation. We assume that higher numbers of AI patents granted translate to higher levels of innovation within the industry, with implications on all jobs. For AI Patent data, we used The Lens (lens.org), a flagship project of Cambia, an independent non-profit social enterprise based in Australia. This project 'seeks to source, merge and link diverse open knowledge sets, including scholarly works and patents' (Lens, 2025). Lens collaborates with renowned organisations like US Patent and Trademark Office (USPTO) and Microsoft Academic as their sources of data points.

Data were filtered out to only provide patent information from the US specifically related to AI and NLP models from any given industry, based on CPC classification codes (see A6). For AI patents, a T-1 lag was used to account for the delayed effect of the business adoption after AI patent releases.

B. Data Omitted

Given the context of the US, minimum wage and educational attainment per occupation were considered in our hypothesis.

Minimum wage may have some implications on the starting wage and therefore the level of wage inequality per skill level. However, we found that US minimum wage policy remained unchanged throughout the time period under consideration (Department of Labor, n.d.*a*). Further research also found only a very small percentage of individuals earning less than the required minimum wage in their occupations (US Bureau of Labor Statistics, 2023). Given these two reasons, minimum wage was omitted from the scope of the regression.

In addition, education levels and education requirement levels per occupation were considered. This is because improvements in education attained by the labor market may have an impact on wage inequality (Kim, 2022). However, we found that there were little to no changes in education levels or requirement per occupation over the years. Besides, our model measures the percentage change of the variables. This variable was almost zero in our regression. It was also therefore removed.

C. Methodology

To calculate wage inequality, we utilize the Theil index, which measures changes in 'entropic distance from the "ideal" egalitarian state of everyone having the same income' (US Census Bureau, 2021).

We aim to look for how AI impacted changes in wage inequality between skill levels. To calculate this, we use a modified Theil Index as shown below:

$$\text{(1)}\quad \text{Inequality\_component}_{i,t} = \frac{1000}{N} \times \left(\frac{w_{i,t}}{\mu_t} \times \operatorname{sgn}\left(\frac{w_{i,t}}{\mu_t}\right) \times \ln\left(\left| \frac{w_{i,t}}{\mu_t} \right| + 1 \right) \right)$$

Where $Inequality\_component_i, t$ is inequality level for group i (e.g.skill levels) at time t. N is the number of skill levels. $w_{i,t}$ is wage growth in percentage terms for each skill level at time t. μ_t is the mean percentage wage growth across all skill groups at time t. $1000/N$ is a per-capita scaling factor, which normalizes results for comparability across datasets. The log transformation downweights extreme values, while remaining sensitive to changes across the entire distribution. A higher outcome in the $Inequality\_component_i, t$ would indicate a higher level of income inequality.

Given this modified Theil Index, we hypothesize that the increase in AI will have an effect on wage inequality between skill groups. We propose a regression model as follows:

$$\begin{aligned} \text{Inequality\_component}_{i,t} = {} & \beta_0 + \beta_1 AIparameters_{i,t} + \beta_2 AIpatents_{i,t} \\ & + \beta_3 skill\_medium_i + \beta_4 skill\_high_i \\ & + \beta_5 (AIparameters \times skill\_medium_i) \\ & + \beta_6 (AIparameters \times skill\_high_i) \\ & + \beta_7 Employment\_change_{i,t} \\ & + \beta_8 GDP\_Growth_{t-1} + \beta_9 PCE\_Growth_{t-1} \\ & + \alpha_i + \epsilon_{i,t} \end{aligned}$$

In order to assess the *change* in AI development and wage inequality, we model all variables as the percentage change between a given year Y_1 and the previous year Y_0 as seen below:

$$Z = \frac{\Delta Y}{Y_0} = \frac{Y_1 - Y_0}{Y_0} = \frac{Y_1}{Y_0} - 1$$

Our core specification is estimated using pooled Ordinary Least Squares (OLS), followed by fixed effects models to control for unobserved heterogeneity. Specifically, we include year-fixed effects to absorb time-specific shocks that affect all occupations uniformly, and occupation fixed effects to account for time-invariant characteristics unique to each occupation. A two-way fixed effects model, incorporating both year and occupation fixed effects, serves as our most comprehensive specification. To address within-group autocorrelation and heteroskedasticity, standard errors are consistently clustered at the occupation level across all estimations.

Furthermore, we introduce the interaction term of AI Parameters and skill levels to assess the heterogeneity of AI's impact in different skill levels, from Low (0, as reference group), Medium (1), and High (2). These terms help to assess if AI development disproportionately affects the labor market, favoring a specific skill level more than the other. Companies decide on the AI models they use based on the AI parameters of the given AI model, ensuring they stay relevant with AI development (Samborska, 2025). In relation to this paper, AI development as measured by AI parameters would positively affect high-skilled laborers and negatively affect low-skilled ones (Ma et al., 2022).

Finally, we combine the left-hand side of the equation (between-group wage inequality measured by the Theil Index) and key explanatory variables (AI Parameters and AI Patents), supplemented with some control factors. These explanatory variables are represented by (β_1) and (β_2)respectively. We also use interaction term between AI Parameters and Skill Level (β_3). Aside from these main variables, we also include changes in employment, GDP growth, and personal consumption expenditures as our control factors. This equation is expected to estimate how improvements in AI restructures wage inequality across different skill levels.

II. Regression Analysis and Interpretations

The results were statistically significant for the AI parameters coefficient (p-value$<$ 0.01). Employment change over the years (p-value$<$ 0.05) was also another primary factor in our prediction. Pearson's correlation coefficient increased to 0.297 in the occupation fixed-effects model relative to the baseline pooled OLS. This means that the model decreases mean-squared error by 29.7% relative to a sample mean estimate and explains 29.7% of the variability in the model.

Using percentage change of Theil index as a dependent variable, we use four model specifications: pooled OLS, year fixed effects, occupation fixed effects, and two-way fixed effects (combined year and occupation fixed effects). Given that the change of Theil index over the years is very marginal, the coefficient started very small. Hence, we scale the model by a factor of 10^5 for easier interpretation. This means that for any 100,000-percent change in the independent variables, the income inequality will increase by the respective coefficients shown above.

The analysis for the effect of AI parameters is broken down into three skill classifications - low, medium, and high skill. The coefficient for low-, medium-, and high-skilled jobs is indicated in the variables AI Parameters, AI Parameters x Medium Skill, and AI Parameters x High Skill respectively. Consistent with our previous assumption, the coefficient on AI parameters is positive and statistically significant at the 1% level. A 100,000 percentage point increase in the percentage change of AI parameters increases the income inequality by 0.12 percentage points, for low-skilled workers on average, holding other variables constant. This effect is also apparent in the occupational fixed effect, with the same significance level. Meanwhile, the coefficients of the interaction terms are not statistically significant. Even so, they show contrasting results to that in low-skilled roles. AI parameters seem to promote wage inequality. In the dummy variables, wage inequality is associated positively in medium-skilled occupations and negatively in high-skilled occupations. This might suggest that the effect of AI parameters on wage inequality is most prominent across low-skilled workers.

For AI patents, we use lagged terms to account for the temporal gap between patent approval and its observable impacts in the workforce. Although the coefficient on AI patents is statistically insignificant, it is negatively linked to wage inequality in both pooled OLS and occupation-fixed effects. This can also potentially imply that AI patents do not directly affect wage inequality in the short run.

Among macroeconomic control variables, changes in employment levels have a statistically significant and positive impact on wage inequality. A percentage point increase in employment change by 100,000 tends to be associated with an increase in the Theil index by 183.724 percentage points. Interestingly, GDP growth and PCE growth are both statistically insignificant in this equation across all the model specifications.

Regarding model fit, incorporating occupation fixed effects greatly enhances the explanatory power of the model, which is shown by the increase of R^2 by 28.5% from 1.2%. This highlights the importance of controlling for structural occupation differences, irrelevant with the development of AI. On the other hand, time-specific shocks, accounted for by year fixed effects, do not change the fitness of the model.

In Table 3, we observe several missing coefficients that are dropped out of the model. This happens because some variables are almost perfectly collinear with others. These pairs include AI Parameters-AI Patents, Medium Skill-High Skill, and GDP Growth-PCE Growth.

Additionally, we assessed the cross-correlation between our estimates and created diagnostic plots for our linear model (see Figure 4). The residuals versus fitted plot has a neither convex nor concave line, indicating a linear relationship. The normal Q-Q plot suggests that the residuals primarily follow a normal distribution. The scale-location plot shows a relatively straight horizontal line with a high concentration of points on the middle and the left, which indicates that our model is resilient towards heteroskedasticity, which is consistent with the results of the Breusch-Pagan test (p-value> 0.05). Finally, the residuals versus leverage plot shows that there are no influential observations on the fitted results, as no observation falls outside the Cook's distance line.

Table 3: Effects of AI on Wage Growth Inequality Components

	Pooled OLS	Year FE	Occupation FE	Two-way FE
AI Parameters	0.120***	-	0.120***	-
	(0.039)		(0.039)	
AI Patents (t-1)	−0.029	-	−0.031	-
	(0.083)		(0.083)	
Medium Skill	1.135	1.135	-	-
	(15.350)	(15.350)		
High Skill	−10.462	−10.462	-	-
	(20.220)	(20.220)		
AI Parameters × Medium Skill	−0.030	−0.030	−0.029	−0.029
	(0.034)	(0.034)	(0.033)	(0.025)
AI Parameters × High Skill	−0.014	−0.014	−0.013	−0.013
	(0.038)	(0.038)	(0.039)	(0.041)
Employment Change (%)	183.724**	183.724**	177.323*	177.323***
	(80.778)	(80.778)	(92.797)	(30.503)
GDP Growth (%, t-1)	−9.386	-	−9.421	-
	(12.087)		(12.164)	
PCE Growth (%, t-1)	8.303	-	8.375	-
	(10.518)		(10.664)	
Num.Obs.	4095	4095	4095	4095
R2	0.012	0.012	0.297	0.297
R2 Adj.	0.009	0.009	0.120	0.120

* p < 0.1, ** p < 0.05, *** p < 0.01
Growth variables expressed in percentage points.

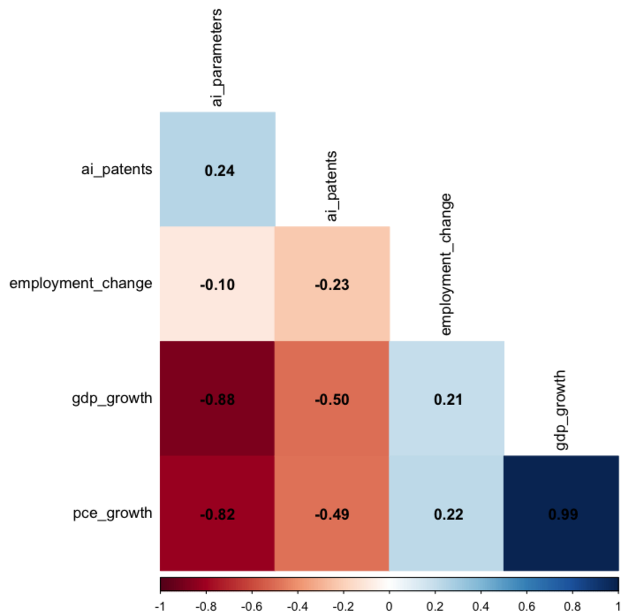

Figure 3. Correlation heat map

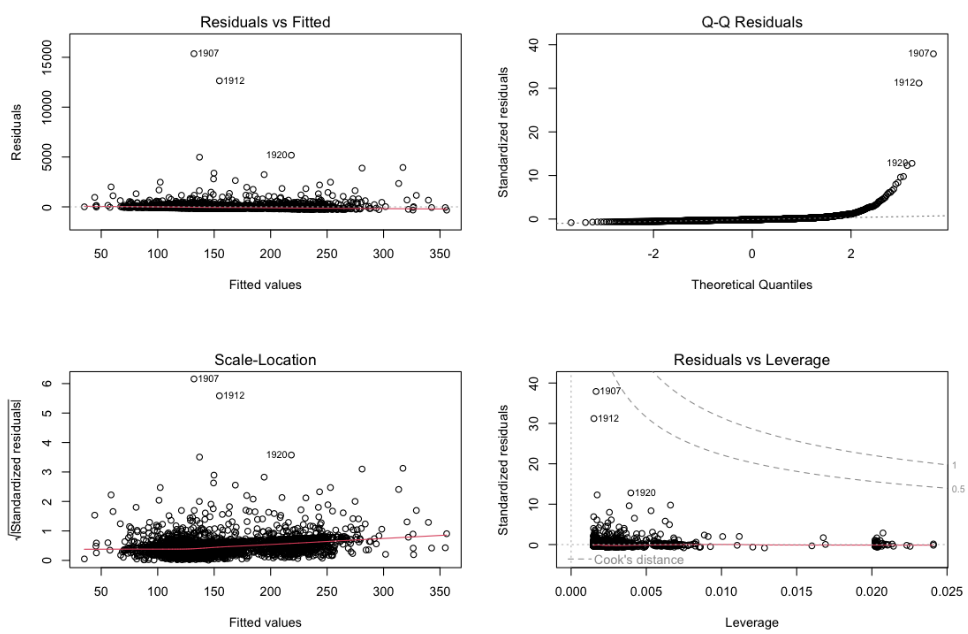

Figure 4. Further Diagnostics Plots

III. Extensions/Robustness Checks/Discussion

Our framework utilizes year fixed effects, also incorporated in the two-way fixed effect, to anticipate time-invariant heterogeneity across occupations and macroeconomic shocks common to all occupations in a given year. However, our assumption is that the relationship between AI advancement and wage inequality is uniform and time-invariant. This might pose a challenge where AI could influence the job market overtime, but gets ignored because of the year fixed effects.

Moreover, the specification relies on Theil index as our dependent variable, wage inequality. Despite having chosen Theil index over other alternatives like Gini coefficient following the claim by World Bank (n.d.), these other metrics could still be applied to our model to compare the overall explanatory power.

Given these shortcomings, to add further value we aim to further expand our study to provide further projections of our model. Specifically, we extrapolate the graph by five years (2025 to 2029) to estimate the effect of AI parameters and AI patent on income inequality. For the prediction, we split the assumptions on three scenarios, conservative, baseline, and optimistic. Conservative AI projections are assumed to have a linear trend. In the baseline segment, the AI growth is assumed to follow quadratic trend. While for the optimistic segment, the AI growth is assumed to be exponential. In all scenarios, only the values of AI parameters are modified according to the presumed growth path, while AI patents are kept linear as it has no statistically significant effect.

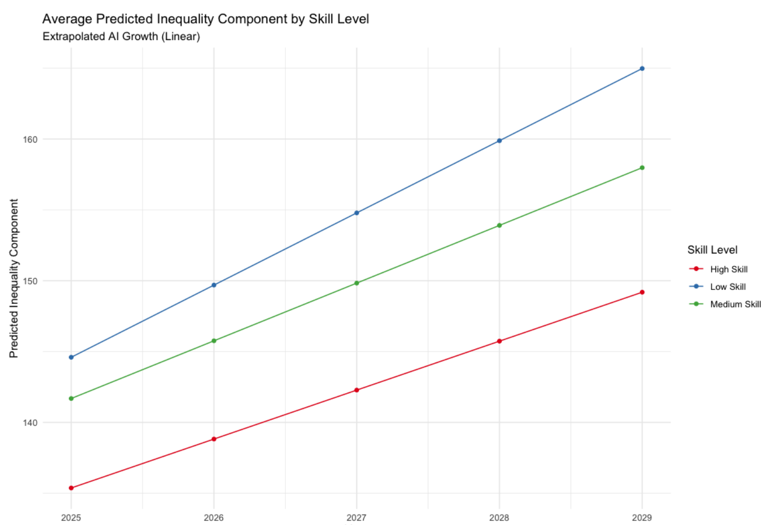

Figure 5. Conservative Forecast

The conservative projection (see Figure 5) extends the graph by assuming the same growth rate of AI parameters, using the mean of growth in AI parameters over the last five years (2020 to 2024). We define linear growth as conservative given that in the previous five years, the AI growth has not been uniform, yet it is still a relevant assumption to include linear trend as our conservative model. In this scenario, high-skilled occupations are expected to have slower growth in income inequality by 3.42% compared to their low-skilled counterparts in the following five years.

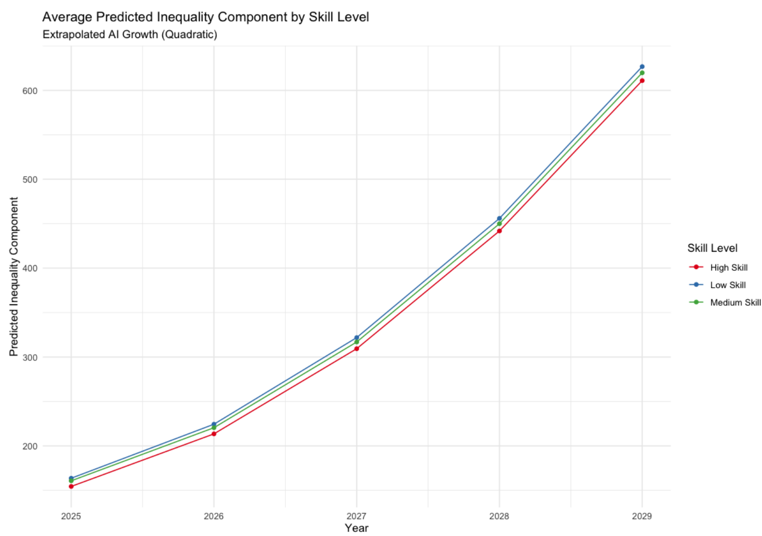

Figure 6. Baseline Forecast

For baseline modeling in Figure 6, we follow the hypothesis that under non-extreme conditions, AI development (represented by AI parameter counts) should continue to grow quadratically. The difference in growth across the three skill-level classifications is predicted to be not that far apart from each other.

Lastly, under the optimistic scenario as in Figure 7, the growth path extends exponentially. This causes all three lines to have marginal distance from one another. At the same time, it means that wage inequality across skill levels may worsen by roughly 2.91 times in the span of five years. That is, if the AI development follows the exponential trend, which is evident based on our previous discussions.

In generating these predictions, we assume a describable trend across the years. However, this might not be the case. AI growth, either AI parameters or AI patents, may have volatile growth in certain years. This is evident in the historical data. AI parameter grew by 355.12% in 2019 but dropped by 41.17% in 2023. Consequently, our analysis focuses more on the endpoint of the timestamp rather than specific value on certain years, which might not be exactly realized in the future.

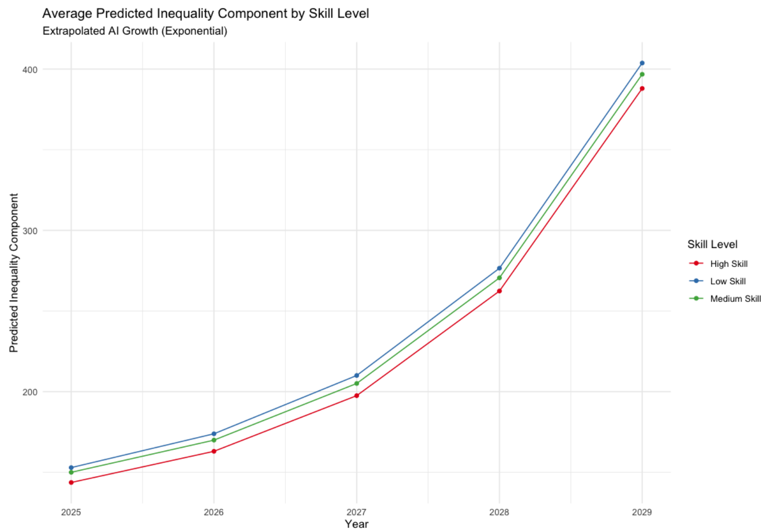

Figure 7. Optimistic Forecast

IV. Literature Review

Current literature surrounding this topic aims to theoretically explain the technological change in wage inequality, or provides the effects of AI on a case-by-case basis. Skill-Biased Technological Change (SBTC) and the task polarization model were the dominating theories we identified to be most relevant. The former explains wage disparities as a result of high-skilled workers being able to 'disproportionately benefit' from technological change, owing to the nature of their work being more abstract-reasoning and communications heavy (Autor, 2022). With the use of computerization, these tasks are complemented, allowing these high-skilled workers to be more productive. This causes an increase in the demand for high-skilled workers, creating a skill premium. In contrast, low-skilled employees whose tasks are not complemented by technological change either face stagnant or declining wages.

The task polarization model follows a similar conclusion, but explains wage disparity as a result of task displacement (Autor, 2022). The theory stipulates that a single job can be broken down into multiple tasks, and routine tasks can be singled out to be automated. Usually, these routine tasks are done by low-skilled workers - and as their tasks are automated, they are consequently displaced. Acemoglu and Restrepo (2022) found that routine-task-intensive jobs faced downward pressure over ensuing decades since automating tasks. Their results showed that 50%-70% of wage changes were tied to automation displacing routine-task workers, implying an adverse impact of technological change on the earnings of less-skilled workers (Acemoglu and Restrepo, 2022).

Other research has involved looking at specific use cases, identifying implications of tailored AI models on workers in a given company. In these cases, results are mixed. A study by Otis et al. (2024) on generative AI's impact on entrepreneurial performance found that well-performing entrepreneurs were able to benefit from the use of AI, but less-performing entrepreneurs did not see the same results. On average, there was no treatment effect of access to the AI model on firm performance (Otis et al., 2024). This is contrary to another study by Brynjolfsson, Li and Raymond (2025), where customer support agents were provided with access to an AI tool and saw a 14% increase in productivity on average, with a disproportionate advantage for novice and low-skilled workers (Brynjolfsson, Li and Raymond, 2025).

Previous research surrounding the topic of AI on skill premium has also been done by Bloom et al. (2024), who explore the impact of technology on the skill premium by proposing a nested constant elasticity of substitution production function and testing three types of capital against it. This study utilized some data from the BLS, showing evidence of AI reducing skill premium, but did not derive specific relationships between AI use and macroeconomic wage data.

To our knowledge, there is no literature to date that analyses the impact of AI on the change in wage inequality between skill levels on a macroeconomic scale. Our research provides novelty through the use of BLS occupational data and their associated income levels, providing realistic analyses of AI's implications on actual changes in income in the US.

V. Conclusion

The implications of AI on the future of labor markets and its constituents remain uncertain. Our paper provides empirical evidence of AI on wage disparity over time on a macroeconomic scale. In our case, we aim to look at the implications of AI parameters and AI patents on a modified Theil Index.

Our results suggest that more complex AI models, measured through the number of parameters, are able to disproportionately help higher-skilled workers, likely due to their ability to leverage AI tools for productivity gains. AI patents, in contrast, showed no significant short term effect, indicating that innovation may not immediately translate into labor market disruption - unlike deployment. The result persists despite our effort to lag the variable by one term. Interactions between AI and skill groups were insignificant, but the coefficient sign confirms that AI's inequality effect is skill-biased, where high-skilled workers capture more gains, while low-skilled workers face stagnation or displacement.

A variety of directions for future research are available from this research. Firstly, while our research is solely focused on skill groups, natural follow up questions would involve AI effects on specific industries. This is mainly because AI development has inconsistent effects on different occupational sectors (Septiandri, Constantinides and Quercia, 2024). One alternative that further studies could consider is to use a more specific set of index systems than AI parameters to measure AI development per industry.

Additionally, non-wage dimensions, like job quality, could capture AI's broader societal implications. Further research can be done to look at AI implications in regions outside of the US, such as China, which is set to peak in AI adoption rates by early 2030 (Goldman Sachs, 2025).

AI is reshaping labor markets, but as our research shows, it is not equal. While AI adoption boosts productivity, its benefits are concentrated among high-skilled workers, exacerbating wage inequality - especially when paired with employment growth in skill-intensive sectors. The capacity of AI may be unknown, but one thing is certain - for a better, more equitable society, AI should be more accessible to all workers - regardless of skill or title.

REFERENCES

Acemoglu, Daron, and Pascual Restrepo. 2022. "Tasks, Automation, and the Rise in US Wage Inequality." *Econometrica*, 90(5): 1973–2016.

Autor, David. 2022. "The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty." National Bureau of Economic Research w30074.

Balakrishnan, T., M. Chui, B. Hall, and N. Henke. 2020. "The State of AI in 2020." *https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2020*, Retrieved April 21, 2025.

Bloom, David E., Klaus Prettner, Jérémy Saadaoui, and M. Veruete. 2024. "Artificial Intelligence and the Skill Premium." National Bureau of Economic Research w32430.

Brynjolfsson, Erik, Daniel Li, and Lindsey Raymond. 2025. "Generative AI at Work." *The Quarterly Journal of Economics*. qjae044.

Bureau of Labor Statistics. 2025. "Occupational Employment and Wage Statistics (OEWS) Tables." *https://www.bls.gov/oes/tables.htm*, Retrieved April 21, 2025.

Department of Labor. n.d.*a*. "Minimum Wage." *https://www.dol.gov/general/topic/wages/minimumwage*, Retrieved April 21, 2025.

Department of Labor. n.d.*b*. "News Release." *https://www.dol.gov/newsroom/releases/osec/osec20241016*, Retrieved April 21, 2025.

Epoch AI. 2024. "Data on Notable AI Models." *https://epoch.ai/data/notable-ai-models*, Retrieved April 21, 2025.

Frey, Carl Benedikt, and Michael Osborne. 2023. "Generative AI and the Future of Work: A Reappraisal." *Brown Journal of World Affairs*, 30: 161–176.

Georgieff, Alexandre. 2024. "Artificial Intelligence and Wage Inequality." OECD Publishing 13.

Goldman Sachs. 2025. "What Advanced AI Means for China's Economic Outlook." *https://www.goldmansachs.com/insights/articles/what-advanced-ai-means-for-chinas-economic-outlook*.

IMF. 2024. "AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity." *https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-\protect\penalty\z@humanity*, Retrieved April 20, 2025.

Jamshidian, Mortaza, and Siavash Jalal. 2010. "Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data." *Psychometrika*, 75(4): 649–674.

Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, and et al. 2020. "Scaling Laws for Neural Language Models." *https://arxiv.org/abs/2001.08361*.

Kim, Hyejin. 2022. "Education, Wage Dynamics, and Wealth Inequality." *Review of Economic Dynamics*, 43: 217–240.

Lee, Kai-Fu. 2018. *AI Superpowers: China, Silicon Valley, and the New World Order.* Harper Business.

Lens. 2025. "What is The Lens." *https://about.lens.org/what/*, Retrieved April 21, 2025.

Ma, Hao, Qiang Gao, Xiang Li, and Yu Zhang. 2022. "AI Development and Employment Skill Structure: A Case Study of China." *Economic Analysis and Policy*, 73: 242–254.

Maher, Timothy, and Niklas Schaffelke. 2023. "Unlocking the AI advantage: Investigating the impact of ai patents on firm earnings and industry dynamics: A comprehensive investigation of the influence of AI patent ownership on corporate financial performance."

Otis, Natalie, Rebecca Clarke, Solene Delecourt, David Holtz, and Rembrand Koning. 2024. "The Uneven Impact of Generative AI on Entrepreneurial Performance." Unpublished manuscript.

Our World in Data. 2025. "Parameters in Notable Artificial Intelligence Systems." *https://ourworldindata.org/grapher/artificial-intelligence-parameter-count*, Retrieved April 21, 2025.

Rosalsky, Greg. 2025. "What America's Top Economists Are Saying About AI and Inequality." *https://www.npr.org/sections/planet-money/2025/01/07/g-s1-41290*, Retrieved April 21, 2025.

Samborska, Victoria. 2025. "Scaling Up: How Increasing Inputs Has Made Artificial Intelligence More Capable." *https://ourworldindata.org*.

Septiandri, A. A., Marios Constantinides, and Daniele Quercia. 2024. "The Potential Impact of AI Innovations on US Occupations." *PNAS Nexus*, 3(9): 320.

Sevilla, Jose, Pablo Villalobos, and Juan Felipe Cerón. 2021. "Parameter Counts in Machine Learning." *https://epoch.ai/blog/parameter-counts*, Retrieved April 22, 2025.

Singla, Anusha, Alex Sukharevsky, Lareina Yee, Michael Chui, and Bill Hall. 2025. "The State of AI: How Organizations Are Rewiring to Capture Value." *https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai*, Retrieved April 21, 2025.

Stanford HAI. n.d.. "The 2025 AI Index Report." *https://hai.stanford.edu/ai-index/2025-ai-index-report*, Retrieved April 21, 2025.

TEDAI San Francisco. n.d.. "What are Parameters in AI?" *https://tedai-sanfrancisco.ted.com/glossary/parameters/*, Retrieved April 24, 2025.

US Bureau of Labor Statistics. 2023. "Wage and Salary Workers Paid Hourly Rates with Earnings at or Below the Prevailing Federal Minimum Wage by Occupation and Industry." *https://www.bls.gov/cps/data/aa2023/cpsaat45.pdf*, Retrieved April 21, 2025.

US Census Bureau. 2021. "Theil Index." *https://www.census.gov/topics/income-poverty/income-inequality/about/metrics/theil-index.html#:~:text=The%20Theil%20index%20is%20a,everyone%20having%20the%20same%20income.*, Retrieved April 21, 2025.

Van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. "MICE: Multivariate imputation by chained equations in R." *Journal of statistical software*, 45: 1–67.

White, Ian R, Patrick Royston, and Angela M Wood. 2011. "Multiple imputation using chained equations: issues and guidance for practice." *Statistics in medicine*, 30(4): 377–399.

World Bank. n.d.. "Glossary | DataBank." *https://databank.worldbank.org/metadataglossary/lac-equity-lab/series/3.0.TheilInd1*, Retrieved April 21, 2025.

Appendix

A1. MICE Imputation

There were several concerns regarding missing chunks of data when performing merges across data tables from 2019-2024. In particular, 3.4% of the overall income data set were missing. On a finer level, 4.3% of income data and 5.3% of employment data per occupation title at a detailed level were missing. As OCC occupation categories became increasingly detailed over time, we performed the merge based on the latest OCC Codes from 2024. This also explains why the majority of missing data points are from earlier recorded entries in the data set.

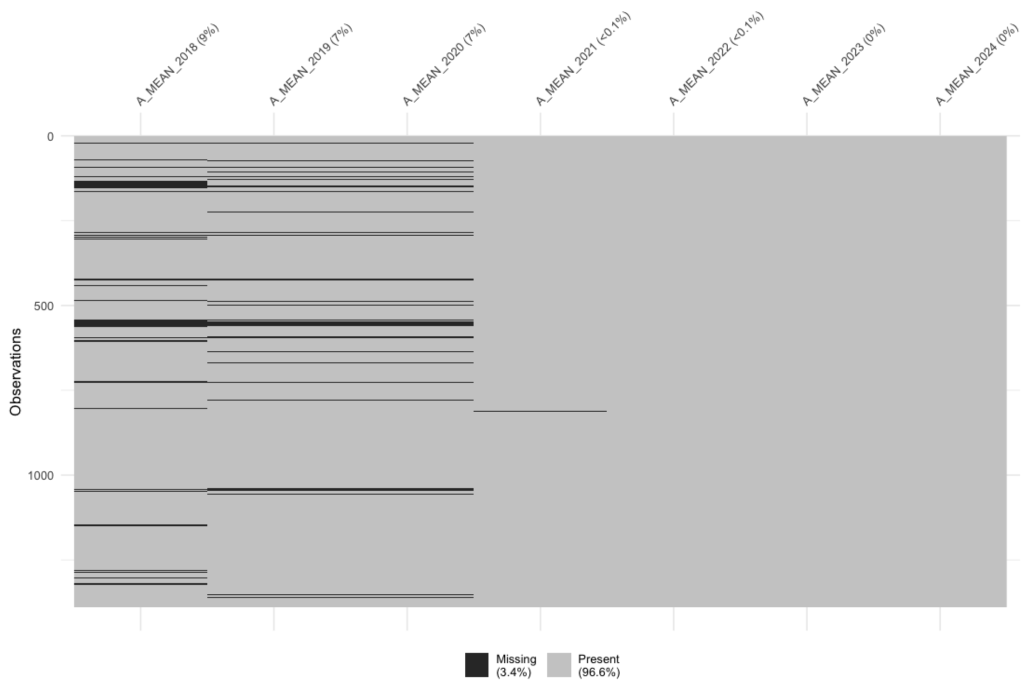

Figure A1. Missing Income Data, Overall

We conducted Little's test on both data sets, which provides strong evidence against the Missing Completely At Random (MCAR) assumption (p-value < 0.001). We then examined the validity of these results by performing a multivariate and univariate normality test. The Mardia test showed that the change in income and the change in employment were not multivariate normal (p-value < 0.001). Furthermore, Anderson-Darling's test concluded that, in fact, the change in income and employment for each and every year was not univariate normal (p-value < 0.001).

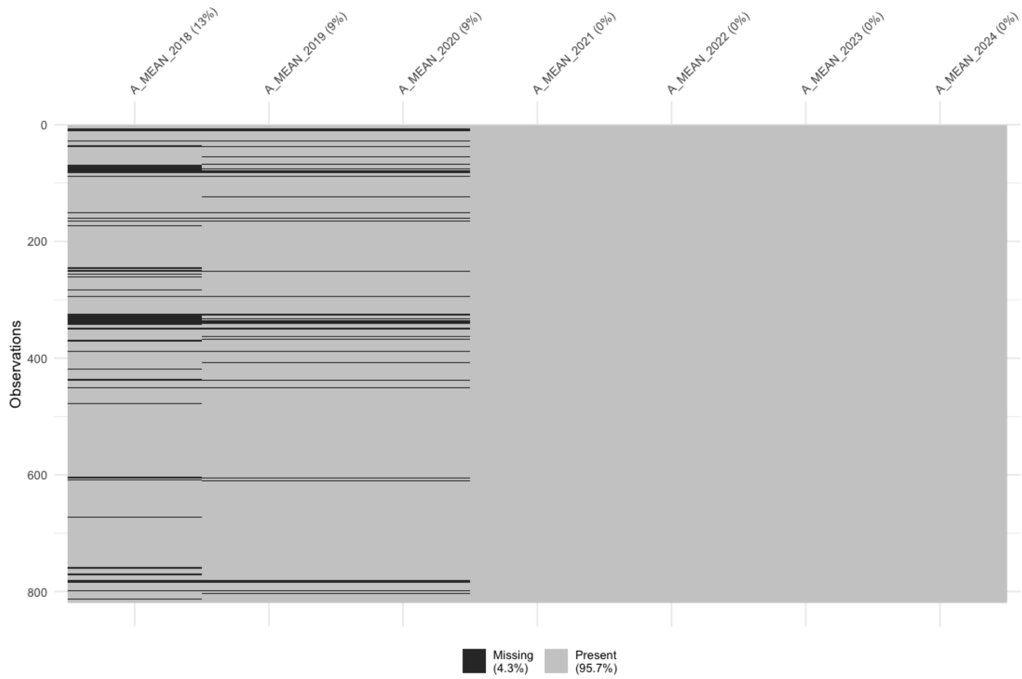

Figure A2. Missing Income Data, Detailed

Despite concerns of violations of the multivariate normality assumption of Little's test, we proceeded to the conclusion that the data are indeed not MCAR. We can clearly visualize missing patterns that differ across time as the above, and we know that OCC standards have increased in detail over time. Furthermore, prior research has shown that Little's test maintains reasonable Type I error rates under moderate departures from normality, particularly with larger sample sizes (Jamshidian and Jalal, 2010).

To further mitigate these concerns, we opt for a more sophisticated method, such as Multiple Imputation by Chained Equations (MICE) with Predictive Mean Matching (PMM). MICE-PMM has been shown to effectively handle longitudinal correlation structures while maintaining robustness to non-normality in the data distribution. The PMM algorithm, in particular, preserves the distributional characteristics of the observed data without assuming normality (Van Buuren and Groothuis-Oudshoorn, 2011; White, Royston and Wood, 2011).

We conducted visual diagnostics for MICE-PMM predictions, assessing the convergence over iterations and a density plot comparing actual values (blue) and imputed values (red). We used earlier data points to predict missing values in the subsequent years. Predictions of both income and employment data converge well. However, density plots for variables over 2018-2019 show that the original

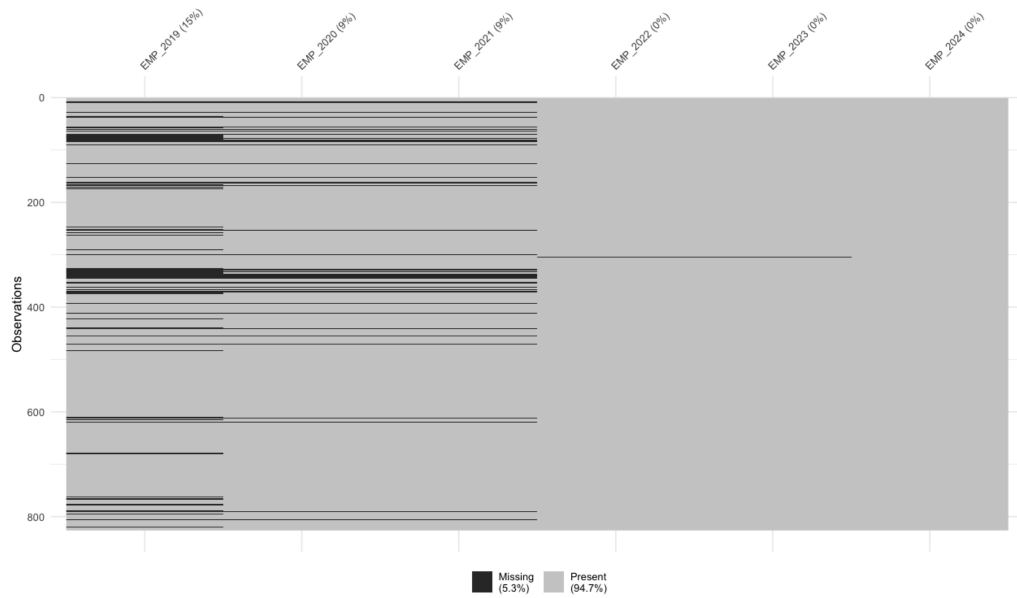

Figure A3. Missing Employment Data, Detailed

distribution was not preserved. This is reasonable, as this is the first period of analysis, and missing values are not imputed with reference to any previous years. Furthermore, there are more missing values relative to later time points. For this reason, we decided to just use data from 2018-2019 for imputation, and we later excluded them from our regression analyses.

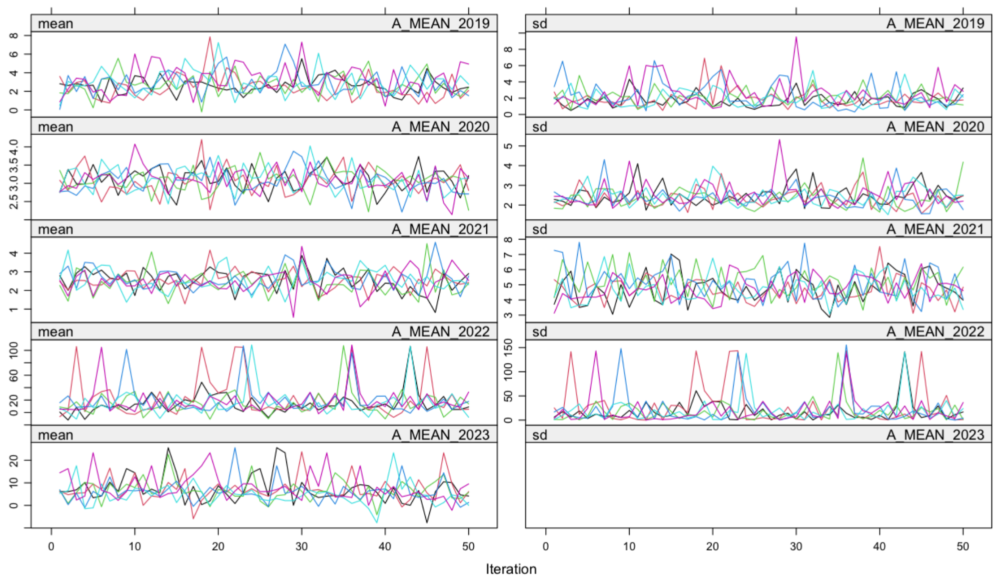

Figure A4. MICE-PMM Convergence, Income

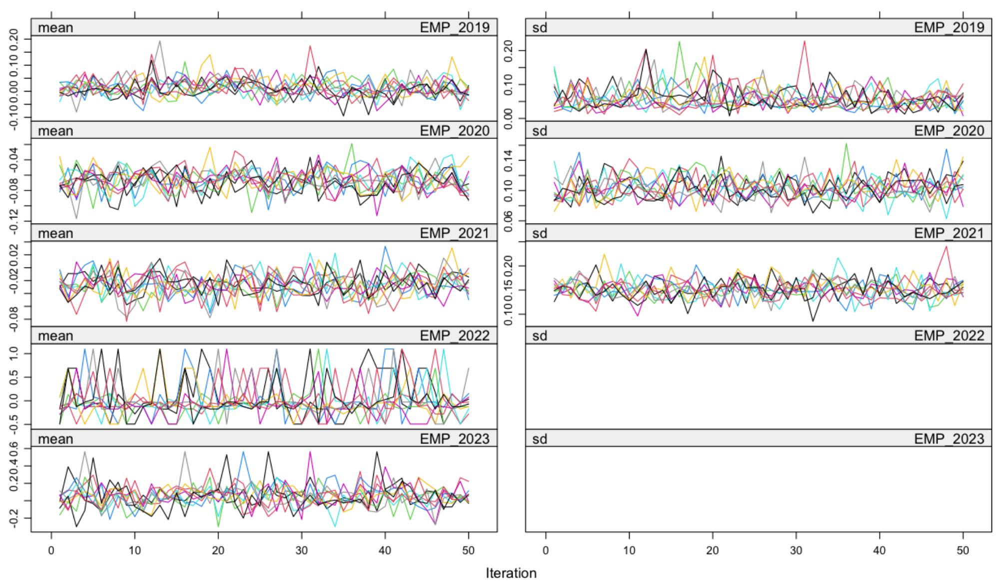

Figure A5. MICE-PMM Convergence, Employment

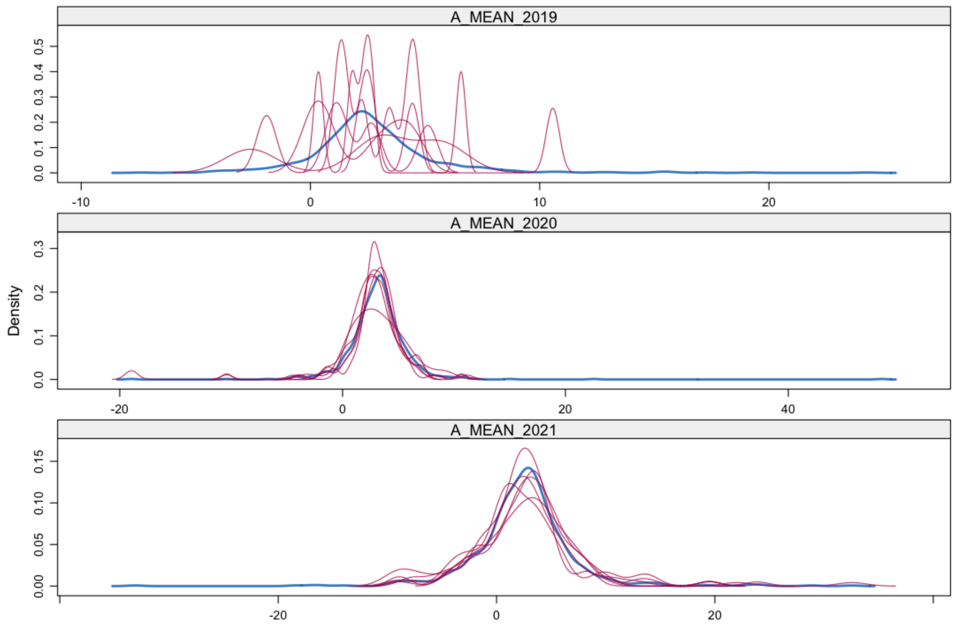

Figure A6. Pre and Post Imputation Density Plots, Income

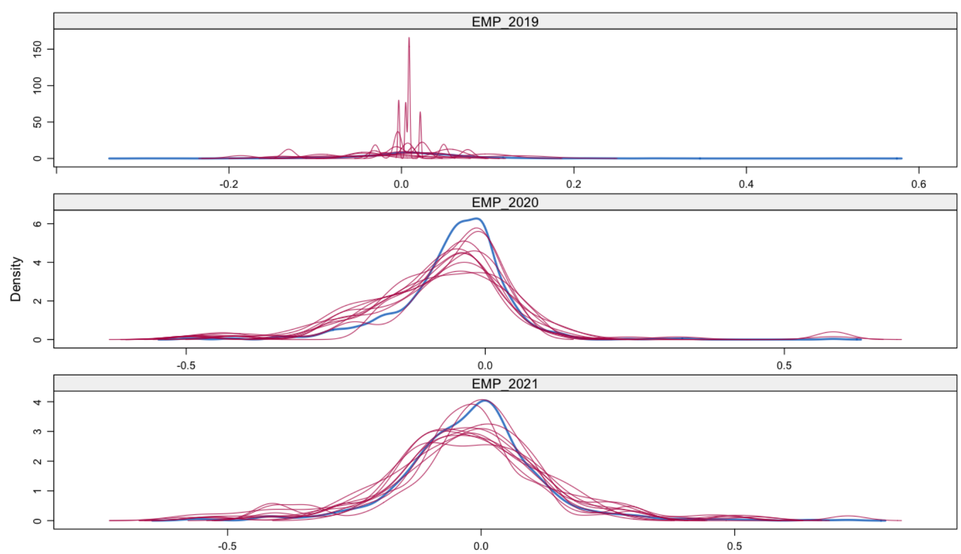

Figure A7. Pre and Post Imputation Density Plots, Employment

A2. ISCO Crosswalk

In the section below, we provide an example of occupations and their corresponding broad skill level.

Broad skill level	**ISCO-08**
Skill levels 3 and 4	1. Managers
	2. Professionals
	3. Technicians and associate professionals
Skill level 2	4. Clerical support workers
	5. Service and sales workers
	6. Skilled agricultural, forestry and fishery workers
	7. Craft and related trades workers
	8. Plant and machine operators, and assemblers
Skill level 1	9. Elementary occupations

Table A1—Broad Skill Levels and ISCO-08

A3. RoBERTa Tiebreaker

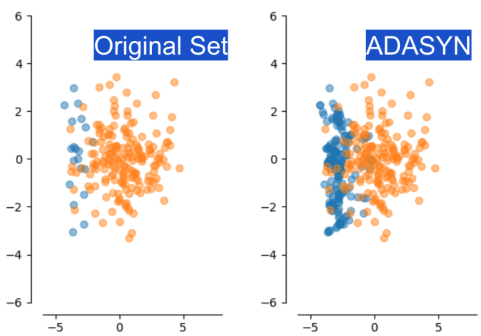

Figure A8. A Simple Example: The Intuitive Effects of ADASYN

Further elaboration on the RoBERTa Tiebreaker can be found below:

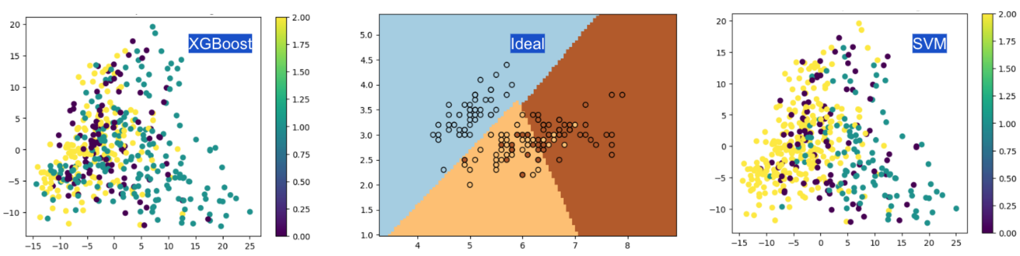

Figure A9. A comparison of ideal, XGBoost, and SVM

Metric	Definition	Low Skill	Medium Skill	High Skill	Weighted Average
Precision	How many of the positively classified are correct?	95%	90%	87%	91%
Recall	How good is it at detecting the positive classes?	100%	82%	90%	91%
F1-Score	Combined measure to gauge both precision and recall	97%	86%	88%	91%
Error Rate	From all classes, positive and negative, how many were predicted incorrectly?	9.32%	**Accuracy**		90.68%

Table A2—Model Performance Metrics Across Skill Levels

Occupations Requiring a Tiebreaker

The following outlines the occupations requiring a tiebreaker in skill level determination, along with the concluding result of the tiebreaker.

	OCC Code	OCC Title	ISCO Skill Level	SVM Result	Final Label
1	11-1021	General and Operations Managers	("2", "1")	2	2
2	11-9021	Construction Managers	("2", "1")	2	2
3	25-3021	Self-Enrichment Teachers	("2", "1")	2	2
4	25-4031	Library Technicians	("2", "1")	2	2
5	27-1012	Craft Artists	("2", "1")	2	2
6	31-9091	Dental Assistants	("2", "1")	2	2
7	31-9096	Veterinary Assistants and Laboratory Animal Caretakers	("2", "1")	2	2
8	31-9099	Healthcare Support Workers, All Other	("2", "1")	2	2
9	33-1012	First-Line Supervisors of Police and Detectives	("2", "1")	2	2
10	33-1021	First-Line Supervisors of Firefighting and Prevention Workers	("2", "1")	2	2
11	33-3051	Police and Sheriff's Patrol Officers	("2", "1")	2	2
12	33-9031	Gambling Surveillance Officers and Investigators	("2", "1")	2	2
13	35-1012	First-Line Supervisors of Food Preparation and Serving Workers	("2", "1")	0	1
14	35-2015	Cooks, Short Order	("1", "0")	0	0
15	35-9011	Dining Room and Cafeteria Attendants and Bartender Helpers	("1", "0")	0	0
16	37-2011	Janitors and Cleaners, Except Maids	("1", "0")	0	0
17	37-2019	Building Cleaning Workers, All Other	("1", "0")	0	0
18	39-1014	First-Line Supervisors of Entertainment and Recreation Workers	("2", "1")	2	2
19	39-1022	First-Line Supervisors of Personal Service Workers	("2", "1")	2	2
20	39-3091	Amusement and Recreation Attendants	("1", "0")	2	1
21	39-4012	Crematory Operators	("1", "0")	2	1
22	39-9099	Personal Care and Service Workers, All Other	("1", "0")	2	1

Table A3—Occupation Classification Results (Part 1: Rows 1–22)

	OCC Code	OCC Title	ISCO Skill Level	SVM Result	Final Label
23	41-1012	First-Line Supervisors of Non-Retail Sales Workers	("2", "1")	1	1
24	41-3041	Travel Agents	("2", "1")	2	2
25	41-9091	Door-to-Door Sales Workers	("1", "0")	1	1
26	41-9099	Sales and Related Workers, All Other	("2", "1")	1	1
27	43-2099	Communications Equipment Operators, All Other	("2", "1")	2	2
28	43-3031	Bookkeeping, Accounting, and Auditing Clerks	("2", "1")	2	2
29	43-4061	Eligibility Interviewers, Government Programs	("2", "1")	2	2
30	43-5021	Couriers and Messengers	("1", "0")	2	1
31	43-5071	Shipping, Receiving, and Inventory Clerks	("2", "1")	1	1
32	45-2011	Agricultural Inspectors	("2", "1")	2	2
33	45-2092	Farmworkers and Laborers, Crop, Nursery, and Greenhouse	("1", "0")	1	1
34	45-2093	Farmworkers, Farm, Ranch, and Aquacultural Animals	("1", "0")	1	1
35	45-4011	Forest and Conservation Workers	("1", "0")	2	1
36	45-4029	Logging Workers, All Other	("1", "0")	0	0
37	47-5099	Extraction Workers, All Other	("1", "0")	0	0
38	49-9099	Installation, Maintenance, and Repair Workers, All Other	("1", "0")	1	1
39	51-4051	Metal-Refining Furnace Operators and Tenders	("2", "1")	1	1
40	51-8099	Plant and System Operators, All Other	("2", "1")	2	2
41	51-9161	Computer Numerically Controlled Tool Operators	("1", "0")	2	1
42	51-9162	CNC Tool Programmers	("2", "1", "0")	2	2
43	53-7065	Stockers and Order Fillers	("1", "0")	0	0
44	53-7199	Material Moving Workers, All Other	("1", "0")	0	0

Table A4—Occupation Classification Results (Part 2: Rows 24–44)

Occupations Requiring a Direct SVM Classification

OCC Code	OCC Title	Final Label
21-1018	Substance Abuse, Behavioral Disorder, and Mental Health Counselors	2
25-2052	Special Education Teachers, Kindergarten and Elementary School	2
25-9045	Teaching Assistants, Except Postsecondary	2
51-2028	Electrical, Electronic, and Electromechanical Assemblers, Except Coil Winders, Tapers, and Finishers	1
53-1047	First-Line Supervisors of Transportation and Material Moving Workers, Except Aircraft Cargo Handling Supervisors	1

Table A5—Occupations Requiring Direct SVM Classification

A4. Cooperative Patent Classification (CPC) Codes used and descriptors

CPC Code	Descriptor
G06N 3/02	Neural networks
G06N 3/08	Learning methods
G06N 5/00	Computing arrangements using knowledge-based models
G06N 7/00	Computing arrangements based on specific mathematical models
G06N 99/00	Subject matter not provided for in other groups (catch-all for emerging AI tech)
G06Q 10/063	AI for business process automation—operations research, analysis, or management
G06N 20/00	Machine learning
G06F 40/00	Handling natural language data
G06V 10/00	Arrangements for image or video recognition or understanding
G16H 50/20	ICT specially adapted for medical diagnosis—for computer-aided diagnosis
G05B 13/02	Adaptive control systems (i.e., systems that adjust themselves for optimal performance)

Table A6—CPC Codes and Descriptors