ITM Web of Conferences **88**, 01034 (2026) https://doi.org/10.1051/itmconf/20268801034
AIBDCS 2026

Artificial Intelligence and Occupational Risk: Evidence from a Longitudinal Panel of Six Economies

Changbin Wang*

College of Computer Science and Technology, China University of Petroleum (East China), QingDao, 266500, China

Abstract. What determines how exposed a worker is to AI substitution, and does that exposure translate directly into lower pay. To address these questions, the paper draws on the Future of Jobs AI Dataset—a panel of 12,343 worker-year observations across 10 occupations in six countries (2015–2035)—and apply a combination of OLS regression with fixed effects, K-means cluster analysis, and two machine learning models. Over the 21-year window, the mean ai_risk_score climbed from 0.316 to 0.467, a 47.9% rise that points toward slow-moving structural change rather than a short-run shock. Regression estimates indicate that the primary technical skill a worker holds matters far more than years of experience or educational attainment: relative to Python specialists, Excel users face an ai_risk_score premium of 0.447 points, and each additional unit of risk is associated with roughly $61,419 less in annual salary. The Random Forest classifier recovers AI risk categories with 83.4% accuracy on held-out data, while a Gradient Boosting model explains 95.5% of salary variance—both results reinforcing the OLS pattern. Cluster analysis groups the ten occupations into three bands whose risk and wage profiles diverge sharply, suggesting that targeted reskilling toward Python, cloud, and deep learning competencies could yield measurable wage gains.

1 Introduction

Consider two workers sitting at adjacent desks in a financial firm. One runs Excel-based reports; the other builds Python pipelines. Their salaries may currently differ by only a modest amount, yet over the next decade the gap is likely to widen considerably—not because of anything they chose, but because of which of their tasks an AI system can most easily replicate. This paper is an attempt to quantify that gap and trace its sources.

Labor market disruption from automation is hardly new, but the current wave has some unusual features. Unlike earlier rounds of mechanization that concentrated on repetitive physical work, AI systems increasingly target analytical and cognitive tasks—the domain of white-collar employment. The World Economic Forum's 2023 Jobs Report projects net job creation of around 12 million roles globally by 2025, yet this aggregate figure conceals losses concentrated in specific occupations and skill groups [1]. Understanding which workers bear the transition costs, and how much of those costs show up in current wages, remains an open empirical question.

The paper addresses it using the Future of Jobs AI Dataset, which tracks 10 occupations across six economies (USA, India, Canada, UK, Germany, Australia) over the period 2015–2035. The dataset's main outcome, ai_risk_score, is a composite index ranging from 0 to 1 that captures the probability of AI-driven substitution based on task structure and technology exposure. Preliminary inspection already hints at large disparities: Business Analysts average a score of 0.75 while AI Researchers sit at 0.15, a difference that reflects task composition far more than any individual characteristic such as education or seniority. Meanwhile, the USA–India salary gap for nominally identical roles stands at 3.5× ($47,629 vs. $13,615), raising the further question of how much of this differential is explained by AI exposure versus other structural factors.

The analysis proceeds in four stages. The paper begins with descriptive statistics to document how risk and wages have evolved across periods, occupations, and countries. The paper then estimates OLS models with country and year fixed effects—separately for ai_risk_score and salary—using HC3 robust standard errors to account for heteroskedasticity. K-means clustering groups occupations into risk-wage profiles, and two machine learning models (Random Forest for risk classification, Gradient Boosting for salary prediction) serve as nonparametric checks on the regression findings. Taken together, the results give reskilling programs a concrete, data-grounded rationale.

2 Literature Review

Existing research establishes several key anchors. Nedelkoska and Quintini estimated that roughly 14% of jobs face high substitution risk, concentrated among younger and less-educated workers [2]. Acemoglu, Autor, Hazell, and Restrepo found that AI-adopting firms reduce non-AI hiring and restructure skill requirements [3]. Webb showed that AI, unlike earlier

*2307020115@s.upc.edu.cn

automation, targets high-skill tasks rather than routine low-wage ones [4], while Felten, Raj, and Seamans constructed the widely used AIOE index linking AI capabilities to occupational requirements [5]. Deming documented growing wage premiums for social and interpersonal skills [6]. More recently, Brynjolfsson, Li, and Raymond found that generative AI disproportionately benefits lower-skilled workers within occupations [7]; Gmyrek, Berg, and Bescond identified clerical work as the most highly exposed broad category globally [8]. Autor et al. further showed that new job titles emerging from technological change systematically complement rather than replicate existing occupational structures, providing a long-run perspective on labor market adaptation [9]. What remains underexplored is how primary technical skill type mediates AI risk over a multi-decade horizon—the gap this paper addresses.

3 Literature Review

3.1 Dataset Overview

The analysis draws on 12,343 complete observations spanning 10 occupations (Data Scientist, Software Engineer, Data Analyst, DevOps Engineer, Cybersecurity Analyst, Cloud Engineer, Business Analyst, AI Researcher, ML Engineer, Product Manager) across six economies from 2015 to 2035. Key variables include ai_risk_score (0.05–0.85), skill_demand_score (60–99), job_openings (1,002–49,998), annual salary in USD (3,875–113,589), primary_skill (nine categories: Python, Java, SQL, Docker, Security, AWS, Excel, Deep Learning, Strategy), education_level (Bachelor's, Master's, PhD), and job_survival_class (0–2). Table 1 presents the descriptive statistics.

Table 1: Descriptive Statistics of Key Variables (N = 12,343)

Variable	Mean	SD	Min	Median	Max
ai_risk_score	0.40	0.19	0.05	0.36	0.85
skill_demand_score	79.45	11.48	60.00	79.00	99.00
job_openings	25,224	14,163	1,002	24,896	49,998
salary (USD)	34,554	18,025	3,875	31,573	113,589

Note: SD = Standard Deviation. No missing values. All observations retained.

3.2 Analytical Methods

Four analytical approaches are employed. Firstly, descriptive statistics are used to characterize risk and wage trends across four five-year periods and across occupations and countries. Secondly, OLS regression models—one for ai_risk_score (M1), one for salary (M2)—are estimated with HC3 robust standard errors and full country and year fixed effects, with Python serving as the reference skill category. Thirdly, K-means clustering is applied to occupation-level means of ai_risk_score, skill_demand_score, and job_openings after standardization, with k=3 selected via silhouette scores (silhouette=0.378). Finally, machine learning extensions are introduced: a Random Forest classifier predicts AI risk category, and a Gradient Boosting regressor predicts salary, both evaluated on held-out test sets and 5-fold cross-validation.

4 Results and Discussion

4.1 Temporal and Cross-Sectional Patterns

Tracking mean ai_risk_score across the four five-year periods reveals a steady upward trajectory: 0.316 in 2015–2019, 0.369 in 2020–2024, 0.416 in 2025–2029, and 0.467 by 2030–2035 (Figure 1). The 47.9% cumulative rise, spread evenly rather than concentrated in a single period, is more consistent with slow structural adjustment than with a discrete technological shock. Wages moved in the same direction—from $28,537 to $40,644 on average—yet this concurrent increase does not contradict the risk story: as Deming notes, wage gains in a period of automation tend to accrue to workers whose competencies complement the new technology, leaving those in substitutable roles behind in relative terms [6].

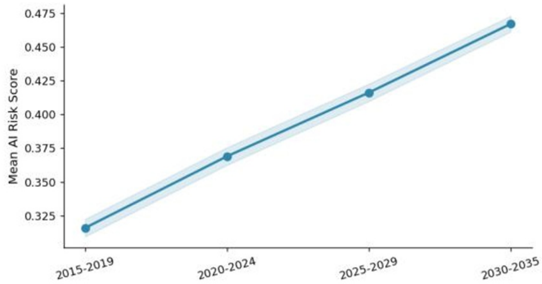

Figure 1: AI Risk Score Trend by Period (95% Confidence Interval), 2015–2035. (Picture credit: Original)

Cross-sectional variation is at least as striking as the time trend. Business Analysts, with a mean risk score of 0.75 and average annual pay of $24,783, occupy the most precarious position in the sample; Data Analysts (0.65, $20,310) are close behind. At the opposite end, AI Researchers score 0.15 and earn $46,128 on average—a risk-adjusted wage premium of over $20,000 relative to Business Analysts (Figure 2). Country-level risk scores are nearly uniform (0.39–0.40 across all six economies), so the 3.5× salary differential between the USA and India ($47,629 vs. $13,615) shown in Figure 3 appears to reflect structural wage-setting differences rather than variation in AI exposure per se [3].

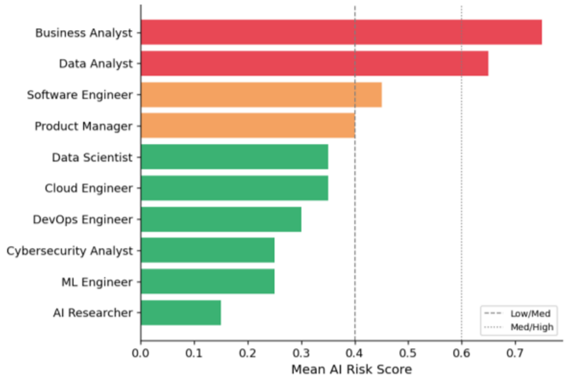

Figure 2: Mean AI Risk Score by Occupation. Green = Low Risk (<0.4), Orange = Medium Risk (0.4–0.6), Red = High Risk (>0.6). (Picture credit: Original)

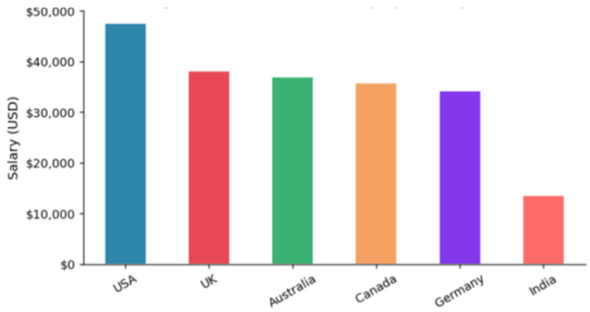

Figure 3: Mean Annual Salary by Country (USD), illustrating the 3.5× USA–India gap. (Picture credit: Original)

4.2 Regression Results

Table 2 presents the central findings. Neither experience nor education has any statistically meaningful relationship with ai_risk_score (all p>0.30), directly contradicting the human capital prediction that more credentialed workers are better insulated [2]. Primary technical skill dominates: Excel workers score 0.447 points above the Python baseline (p<.001), SQL specialists 0.346 points above, and Java specialists 0.144 points above; model R^2 reaches 0.882. The skill coefficient pattern is clearly visualized in Figure 4, which shows Deep Learning as the most protective skill (−0.153 vs. Python baseline) while Excel carries the highest penalty.

In the salary model (M2, R^2=0.869), experience adds $12,766 per step (p<.001) while education effects are negligible. Most consequentially, each unit increase in ai_risk_score carries a salary penalty of $61,419 (p<.001), net of all other covariates—workers in high-risk roles thus face both lower current wages and growing future exposure.

Table 2: OLS Regression Results — ai_risk_score (M1) and Salary (M2)

Predictor	**M1 Coef.**	**M1 p**	**M2 Coef.**	**M2 p**	**Sig.**
Experience level (0–2)	−0.000	.778	+12,766	<.001	ns / ***
Education: Master's	+0.000	.723	−107	.458	ns / ns
Education: PhD	+0.001	.358	−225	.119	ns / ns
Skill: SQL (ref. Python)	+0.346	<.001	−38,575	<.001	***
Skill: Java	+0.144	<.001	−16,161	<.001	***
Skill: Excel	+0.447	<.001	−41,697	<.001	***
Skill: Deep Learning	−0.153	<.001	+16,352	<.001	***
ai_risk_score	—	—	−61,419	<.001	***
Country & Year FE	Yes		Yes		
R^2	0.882		0.869		

Note: M1 DV = ai_risk_score; M2 DV = annual salary (USD). HC3 robust standard errors. N = 12,343. *** p < .001; ns = not significant.

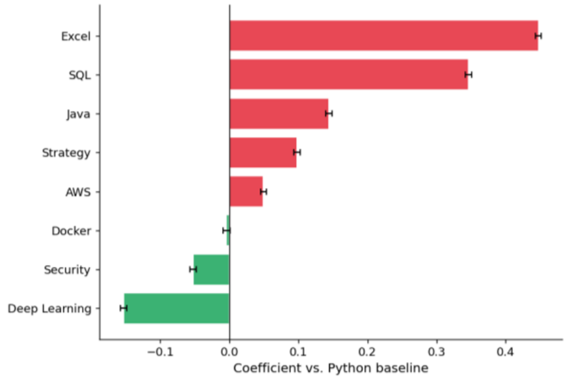

Figure 4: Primary Skill Coefficients vs. Python Baseline (M1), with 95% Confidence Intervals. Positive values indicate higher AI risk than Python workers. (Picture credit: Original)

4.3 Machine Learning Extensions

To complement the OLS analysis, two machine learning models provide additional predictive validity. A Random Forest classifier predicts AI risk category (Low/Medium/High) with 83.4% test accuracy and 5-fold cross-validation mean of 83.0% (±0.4%), substantially above a naive baseline. Feature importance analysis (Figure 5) confirms primary_skill as the dominant predictor (importance=0.615), followed by salary (0.202), with education contributing minimally (0.010)—converging with the OLS finding that skill type, not credentials, determines exposure. The confusion matrix (Figure 6) shows particularly high precision for High Risk (93%) and lower recall for Low Risk (63%), suggesting boundary ambiguity in the medium-risk zone.

A Gradient Boosting regressor predicts annual salary with MAE=$2,641 and R^2=0.955 (Figure 7), substantially outperforming the OLS salary model

(R^2=0.869). The tighter prediction band in Figure 7 validates the observed wage structure and confirms that the salary penalty associated with ai_risk_score is not an OLS artifact but a robust, nonlinearly-confirmed pattern.

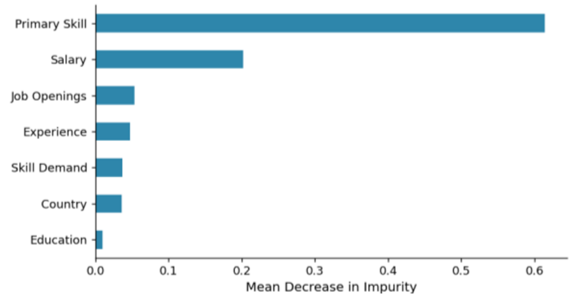

Figure 5: Random Forest Feature Importances for AI Risk Category Prediction. Primary Skill accounts for 61.5% of predictive power. (Picture credit: Original)

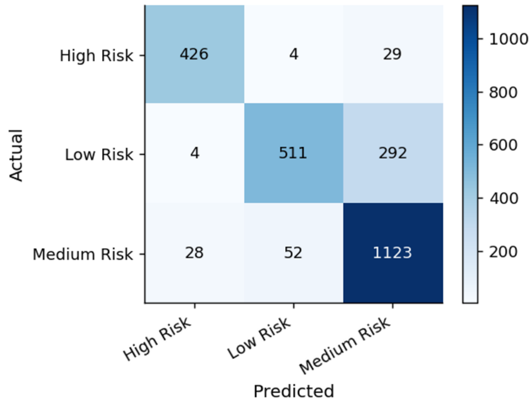

Figure 6: Confusion Matrix for Random Forest Risk Category Classifier (Test Set, N = 2,469). (Picture credit: Original)

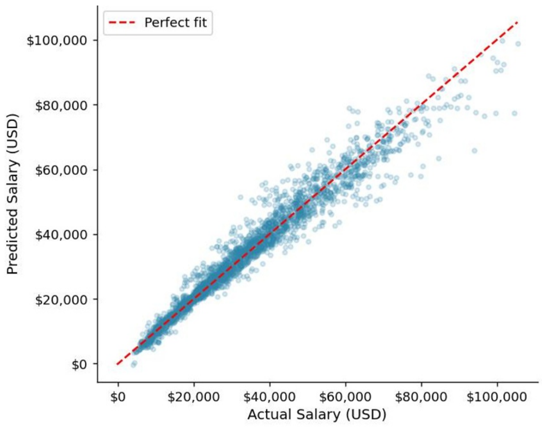

Figure 7: Gradient Boosting — Predicted vs. Actual Salary (R^2=0.955, MAE=$2,641). (Picture credit: Original)

4.4 Occupational Clusters and Policy Implications

The K-means solution (k=3, silhouette=0.378) yields three clusters with distinct profiles. Cluster A (Business Analyst, Data Analyst, Software Engineer) combines the highest risk (mean=0.62) with the lowest salaries (mean=$25,500), reflecting task structures dominated by structured, codifiable work. Cluster B (Cybersecurity Analyst, Data Scientist) occupies a middle position (risk=0.30; salary=$32,933). Cluster C (AI Researcher, ML Engineer, Cloud Engineer, DevOps Engineer, Product Manager) sits at the favorable end: low risk and strong salaries (mean=$40,976), consistent with the growing premium on AI-complementary work [5][6]. The silhouette analysis confirming k=3 is presented in Figure 8, and the cluster scatter is shown in Figure 9.

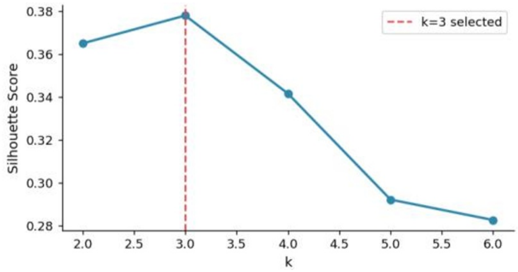

Figure 8: Silhouette Scores for k = 2–6. k=3 selected (score=0.378). (Picture credit: Original)

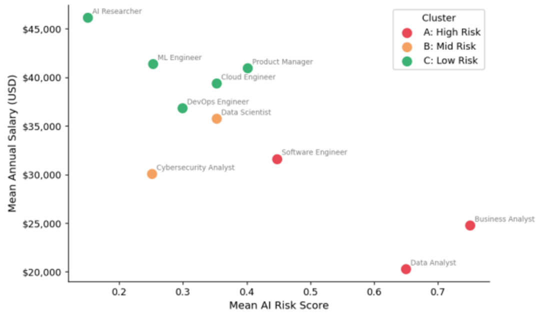

Figure 9: Occupational Clusters — Mean AI Risk Score vs. Mean Annual Salary. Each point represents one occupation. (Picture credit: Original)

For Cluster A workers, the policy implication is direct: transitioning from Excel or SQL to Python, AWS, or Deep Learning would reduce AI exposure and, given the $61,419 salary penalty per risk unit, likely produce immediate wage gains. For Cluster C, the bottleneck is supply rather than demand; expanding advanced technical education and streamlining credential pathways would address the mismatch more effectively than demand-side interventions [2][7][10].

5 Conclusion

The paper sets out to determine whether AI-related occupational risk in a multi-economy panel is driven primarily by individual human capital or by the technical skill a worker happens to hold—and whether that risk registers in current wages. The answer on both counts is unambiguous. Across 12,343 observations and 21 years, the skill an individual uses at work predicts ai_risk_score far better than education level or years of experience, neither of which reaches statistical significance in the preferred specification. Excel specialists, for instance, carry a risk premium of 0.447 points relative to Python workers—a gap large enough to shift them from the medium-risk to the high-risk cluster. That gap has a direct wage correlate: the estimate of $61,419 in lost annual salary per unit of ai_risk_score implies that the highest-risk workers in the sample forgo a substantial portion of the wages earned by their lowest-risk counterparts, holding country and year fixed.

The machine learning results strengthen rather than complicate this picture. A Random Forest model classifies workers into risk categories with 83.4% accuracy, and primary skill alone accounts for 61.5% of that model's predictive power—a share far exceeding salary, job openings, or country of employment. Gradient Boosting recovers 95.5% of salary variance, validating the wage structure estimated by OLS through a fully nonparametric lens. Taken together, these findings give workforce development programs a concrete empirical anchor: the decision to transition from legacy spreadsheet skills toward Python, cloud platforms, or deep learning is not merely speculative career advice but is grounded in measurable differences in both risk exposure and observed compensation.

Several limitations bear mentioning. The panel extends to 2035, meaning that a portion of the observations are forward-looking projections rather than realized outcomes, and the dataset is synthetically generated rather than drawn from administrative records. Future work incorporating matched employer-employee data across multiple countries would allow sharper identification of the causal wage effect and could test whether the skill-risk gradient identified here holds once firm-level heterogeneity is controlled for. As AI capabilities continue to evolve, periodic re-estimation of these relationships will also be necessary.

References

1. World Economic Forum. (2023). The Future of Jobs Report 2023. World Economic Forum, Geneva.
2. Nedelkoska, L., and Quintini, G. (2018). Automation, Skills Use and Training. OECD Social, Employment and Migration Working Papers, 202.
3. Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. Journal of Labor Economics, 40(S1), S293-S340.
4. Webb, M. (2019). The Impact of Artificial Intelligence on the Labor Market. Stanford University Working Paper.
5. Felten, E., Raj, M., and Seamans, R. (2021). Occupational, industry, and geographic exposure to AI. Strategic Management Journal, 42(12), 2195-2217.
6. Deming, D.J. (2017). The growing importance of social skills in the labor market. Quarterly Journal of Economics, 132(4), 1593-1640.
7. Brynjolfsson, E., Li, D., and Raymond, L. (2023). Generative AI at Work. NBER Working Paper, 31161.
8. Gmyrek, P., Berg, J., and Bescond, D. (2023). Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality. ILO Working Paper, 96.
9. Autor, D., Chin, C., Salomons, A., and Seegmiller, B. (2024). New frontiers: The origins and content of new work, 1940–2018. Quarterly Journal of Economics, 139(3), 1399-1465.
10. Acemoglu, D., and Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. Journal of Political Economy, 128(6), 2188-2244.