Proceedings of ICBioMed 2025 Symposium: AI for Healthcare: Advanced Medical Data Analytics and Smart Rehabilitation
DOI: 10.54254/2753-8818/2025.AU24404

Modeling and Forecasting the Dynamic Impact of Artificial Intelligence on the US Labor Market: Heterogeneity Analysis Based on Machine Learning and Mixed Effects Models (2018-2024)

Jingping Liang[1*†], Shiyao Qin[2†], Shuhao Zhang[2†]

[1]*Guangdong Experimental High School, Guangzhou, China*
[2]*Shandong Experimental High School, Jinan, China*
**Corresponding Author. Email: estella_liang@163.com*
[†]*These authors contributed equally to this work and should be considered co-first authors.*

Abstract. This study systematically evaluates the dynamic impact of artificial intelligence (AI) on the US labor market from 2018 to 2024, with a focus on industry-level heterogeneity. Leveraging machine learning and Mixed-effects Modeling, we integrate random forests, Generalized Additive Models (GAM), and Mixed-effects frameworks to quantify nonlinear relationships and sector-specific sensitivities. Key findings reveal that AI market value (variable importance = 0.553 ± 0.028) dominates AI adoption (0.447 ± 0.032) in predicting net unemployment, reflecting the amplified economic valuation effects. GAM analyses identify a threshold effect: AI market value exhibits significant marginal impacts below 0.4 (slope = 0.28 ± 0.05, $p < 1 \times 10^{-16}$), plateauing at higher values. Mixed-effects models further uncover stark industry disparities, with tourism demonstrating robust positive sensitivity to AI development (random slope = 109.37, $p < 0.001$), contrasting with negative responses in manufacturing (-12.99) and services (-25.43). Methodologically, the study highlights the utility of nonlinear modeling (pseudo $R^2 = 0.384$) but underscores limitations from small sample constraints (n = 8), including wide confidence intervals (AICc = 24.62 vs. AIC = 22.24) and residual right-skewness (skewness = 1.02 ± 0.18). By bridging methodological rigor with actionable insights, this work advances the empirical understanding of AI's heterogeneous labor market impacts.

Keywords: component: Artificial Intelligence (AI), Unemployment, Industry Heterogeneity, Machine Learning, Mixed Effects Models, Generalized Additive Models (GAM), Nonlinear Relationships, Industry Sensitivity, Random Forest

1. Introduction

With the rapid development of artificial intelligence (AI) technology, it has had a profound impact on the production model and occupational structure of the US labor market. This impact has a "double-edged sword" effect: on the one hand, it has enhanced production efficiency and created new job

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.AU24404

opportunities; on the other hand, it has led to the disappearance of some jobs and adjustments in the employment structure. Therefore, systematically quantifying the dynamic impact of AI on the job market is of great significance. This study aims to reveal the differentiated impact of the diffusion of AI technology on employment in different industries and provide empirical support for policymakers to optimize labor force transformation strategies.

1.1. Current research status and limitations

1.1.1. Limitations of research methods

At present, certain progress has been made in the research on the impact of artificial intelligence on the job market, but there are still some limitations and deficiencies. Some studies use a capability-based classification method to assess the risk of occupations being affected by artificial intelligence. Frey & Osborne (2017) only evaluated occupational risks based on nine capability classifications (manual dexterity, finger dexterity, perceptual speed, creativity, fine manipulation, mathematical reasoning, spatial orientation, written comprehension, and oral comprehension) [1]. However, this approach may not fully capture the complex impact of artificial intelligence on occupations, as the diversity and complexity of occupations in the United States exceed the scope of these nine capabilities. Additionally, Brynjolfsson & McAfee (2023) explored the relationship between AI-driven productivity improvements and labor market disruptions from a global perspective [2]. However, their research may have limitations in data sources, including subjective biases in survey data from industries like finance, technology, and healthcare in the United States, which could affect the accuracy of the research results.

1.1.2. Limitations of research scope

Most existing studies focus on macro-level analysis, with relatively few studies on the heterogeneous impacts across different industries. For example, Acemoglu & Restrepo (2020) pointed out that the impact of artificial intelligence and automation on work is a complex process involving the interaction of multiple factors [3]. However, they mainly focused on the overall impact and concentrated on the macro employment market in the United States, without delving into the differences among different industries. Moreover, most studies have not adequately considered the temporal dynamics of the diffusion of artificial intelligence technology and the differences in the speed and manner in which different industries adapt to it.

1.2. Research questions

To address these limitations, this study proposes the following research questions:

What is the mathematical relationship between AI adoption rates, market valuation, and net employment changes across industries?

While prior studies acknowledge AI's dual effects, the nonlinear interplay between economic valuation (e.g., AI market value) and adoption rates remains underexplored. This question quantifies how these drivers jointly shape employment outcomes, particularly under threshold effects.How do industry-specific sensitivities to automation risks explain heterogeneity in employment dynamics? Existing macro-level analyses overlook sectoral disparities in technological adaptability. By integrating mixed-effects modeling, this question disentangles universal trends (e.g., AI adoption's fixed effects) from industry-specific vulnerabilities (e.g., random slopes for manufacturing vs. tourism), identifying sectors most resilient or susceptible to AI disruptions.

DOI: 10.54254/2753-8818/2025.AU24404

1.3. Research methodology and contributions

This study employs multiple methods to model and predict the impact of artificial intelligence on the US job market.

Firstly, the random forest model, proposed by Hastie & Tibshirani (1990) [4] as an ensemble learning method, is utilized to assess the global importance of variables by constructing multiple decision trees.

Secondly, the Generalized Additive Model (GAM) is employed, which captures complex nonlinear relationships between variables and responses through smoothing functions.

Moreover, the mixed-effects model, given the limitations of existing studies, such as Rey and Osborne's assessment based on only nine ability classifications; Brynjolfsson and McAfee's study with limited data sources [5]; and Acemoglu and Restrepo's focus on the overall impact without detailed analysis of different industries, fills the research gap: specifically, because it can simultaneously consider fixed and random effects, it can more comprehensively reveal the dynamic impact of artificial intelligence on the US job market. Also, the fixed-effects model can quantify the overall trend impact of artificial intelligence adoption rates and industry automation risks on employment; while the random-effects model can analyze the sensitivity differences of different industries to artificial intelligence, which helps to more accurately capture the heterogeneous impacts among different industries.

The significance of industry heterogeneity lies in the fact that different industries have significant differences in their adaptation speed, methods, and impact degrees to artificial intelligence technology. For instance, some industries may achieve higher production efficiency and profit growth due to the application of artificial intelligence technology, while others may face job losses or business model disruptions. Therefore, considering industry heterogeneity can fill the gap in existing research on the differentiated impacts on different industries and provide more targeted labor force transformation strategy suggestions for policymakers.

The data for this study is sourced from relevant statistics during the period from 2018 to 2024, covering key variables such as the market value of artificial intelligence, the adoption rate of artificial intelligence, industry automation risks, and employment situations in different industries. Through these data, the dynamic impact of artificial intelligence on the US job market can be modeled and predicted.

2. Methodology

2.1. Data prepossessing

In this study, data preprocessing is a fundamental step in the modeling and analysis, aiming to ensure the quality and applicability of the data. The following are the specific steps of data preprocessing:

Data cleaning

Missing value processing: filling or deleting missing values to ensure data integrity.

Outlier detection: Identify and handle outliers through box-and-line plots and Z-score methods.

Data Normalization

Normalization Processing: Normalize the data to the [0, 1] interval to eliminate the effect of scale differences on the model.

Cube Root Transformation: Perform cube root transformation on the data to reduce the skewed distribution of the data and ensure the stability of the data.

Industry Classification Optimization

DOI: 10.54254/2753-8818/2025.AU24404

The current study categorizes industries into five categories (agriculture, industry, services, tourism, and manufacturing), but the categorization is rough. Future research could further refine the industry classification to the NAICS subcode level to capture industry heterogeneity more comprehensively.

Definition of variables

Table 1 lists the variables used in this study and their descriptions:

Table 1: Variables and their descriptions

Variables	Description
AI Software Revenue (in Billions)	Annual revenue generated from AI software
Global AI Market Value (in Billions)	The global market value of AI
AI Adoption (%)	Percentage of organizations adopting AI
Organizations Using AI (%)	Percentage of organizations currently using AI
Organizations Planning to Implement AI (%)	Percentage of organizations planning to adopt AI
Global Expectation for AI Adoption (%)	Global expectations for AI adoption
Estimated Jobs Eliminated by AI (millions)	Estimated number of jobs eliminated globally due to AI
Estimated New Jobs Created by AI (millions)	Estimated number of new jobs created globally due to AI
Net Job Loss in the US	Estimated net job loss in the U.S. due to AI
Organizations Believing AI Provides Competitive Edge (%)	Percentage of organizations that think AI gives them a competitive edge
Companies Prioritizing AI in Strategy (%)	Percentage of companies prioritizing AI in their strategy
Estimated Revenue Increase from AI (trillions USD)	Estimated increase in company revenue due to AI
Marketers Believing AI Improves Email Revenue (%)	Percentage of marketers who believe AI enhances email revenue
Expected Increase in Employee Productivity Due to AI (%)	Expected percentage increase in employee productivity due to AI
Americans Using Voice Assistants (%)	Percentage of Americans using voice assistants
Digital Voice Assistants (billions of devices)	Number of digital voice assistant devices (in billions)
Medical Professionals Using AI for Diagnosis (%)	Percentage of medical professionals using AI for diagnosis
AI Contribution to Healthcare (in Billions)	Economic contribution of AI to the healthcare industry (in billions)
Jobs at High Risk of Automation - Transportation & Storage (%)	Percentage of jobs at high risk of automation in the transportation and storage sector
Jobs at High Risk of Automation - Wholesale & Retail Trade (%)	Percentage of jobs at high risk of automation in the wholesale and retail trade sector
Jobs at High Risk of Automation - Manufacturing (%)	Percentage of jobs at high risk of automation in the manufacturing sector
Agricultural Profit Value (in $100 million)	Profit amount in the agricultural sector (in units of $100 million)
Industrial Profit Value (in $100 million)	Profit amount in the industrial sector (in units of $100 million)
Service Profit Value (in $100 million)	Profit amount in the service sector (in units of $100 million)
Touristic Profit Value (in $100 million)	Profit amount in the tourism sector (in units of $100 million)
Manufactural Profit Value (in $100 million)	Profit amount in the manufacturing sector (in units of $100 million)
Agricultural Overall Industry Scale (in $100 million)	Total value of the agricultural industry (in units of $100 million)
Industrial Overall Industry Scale (in $100 million)	Total value of the industrial sector (in units of $100 million)
Service Overall Industry Scale (in $100 million)	Total value of the service sector (in units of $100 million)
Touristic Overall Industry Scale (in $100 million)	Total value of the tourism sector (in units of $100 million)
Manufactural Overall Industry Scale (in $100 million)	Total value of the manufacturing sector (in units of $100 million)
Industrial AI Development Value	Value of AI-related advancements within the industrial sector

DOI: 10.54254/2753-8818/2025.AU24404

The data for this study comes from the publicly available U.S. job market dataset on the Kaggle platform[6] as well as the Bureau of Labor Statistics (BLS), covering a total of eight annual observations from 2018 to 2024, with each year corresponding to an independent data point. The data variables include AI Adoption, AI Market Value, Industry Automation Risk, and Net Unemployment in an annual aggregated form, aiming to analyze the dynamic impact of AI technology on the labor market and its industry heterogeneity.

2.2. Model

2.2.1. Random forest

Description

Random Forest is an integrated learning algorithm that improves model stability and accuracy by constructing multiple decision trees and combining their predictions. The core idea of Random Forest is to use Bootstrap Aggregation and "Random Feature Selection" to reduce overfitting and improve the generalization ability of the model.

In this study, Random Forest is used to evaluate the impact of two key variables, AI Adoption and AI Market Value, on the net number of unemployed. Random Forest's variable importance analysis allows for the quantification of the extent to which each variable contributes to the model's predictions, thus identifying the factors that have a greater impact on the number of net unemployed.

Model Formulation

The random forest model consists of multiple decision trees, and each decision tree is constructed as follows:

Data sampling:

Random samples (with put-backs) are taken from the original dataset by self-sampling method (Bootstrap) to form multiple training subsets.

Feature Selection:

At each node split, randomly select a portion of features as candidate features.

Decision Tree Construction:

For classification problems, the Gini coefficient or information gain is used; for regression problems, the mean square error (MSE) is used to select the optimal split point.

Tree complexity is controlled by limiting the maximum depth or minimum number of sample splits.

Integrated prediction:

The classification problem integrates the results by majority voting method and the regression problem outputs the final results by averaging the predicted values.

In this study, the Random Forest Model is formulated as follows:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(x)$$

where:

$\hat{y}$ is the predicted value of the model for the target variable

T is the number of decision trees;

$f_t(x)$ is the predicted value of the input feature x by the t th decision tree

2.2.2. Generalized Additive Model

Description

Generalized Additive Model, GAM is a statistical model for capturing nonlinear relationships between variables. Unlike traditional linear models, GAM better handles complex nonlinear patterns by introducing a smoothing function (e.g., a spline function) to model the relationship between features and response variables.

DOI: 10.54254/2753-8818/2025.AU24404

In this study, GAM is used to analyze the nonlinear effects of two key variables (AI Adoption and AI Market Value) on net unemployment. Smoothing functions and significance tests allow the marginal effect of each feature on the number of net unemployed to be assessed.

Model Formulation

In this study, the Generalized Additive Model is formulated as follows:

$Y = \alpha + f_1 + f_2 + \epsilon$

where,

Y is the net loss of jobs in the job market due to AI technology adoption.

X_1 is the penetration level of AI technology in each industry or region

X_2 is Industry Automation Risk which is the probability of jobs being replaced by AI in each industry

α is the intercept term which represents the baseline value of the response variable when predictors are zero.

f_1 , f_2 is the smoothing functions which capture nonlinear relationships between predictors and the response variable.

ϵ is the error term that accounts for unexplained variability or noise in the model.

$L = p(y_i \mid \mu_i) - \lambda^2 dx$

where,

log-likelihood term

Log-likelihood function: $p(y_i \mu_i)$

This item measures the goodness of fit of the model to the observation y_i .

y_i is the ith observation.

μ_i is the mean of the model predictions, related to y_i .

p is the conditional probability density function that expresses the probability that observation yi will occur given the predicted mean μ_i .

The log-likelihood function obtains an overall likelihood measure by taking the logarithm of the conditional probabilities of all observations and summing them.

Regularization term

Regularization parameter: λ

λ is the regularization parameter, used to control the complexity of the model.

Larger values of λ lead to smoother functions, which reduces the risk of overfitting.

A larger value of λ allows for more complex functions, which may fit the data better, but may reduce the risk of overfitting.

Smaller values of λ allow for more complex functions that may fit the data better, but may also increase the risk of overfitting.

Smoothness penalty term: $\sum_{j=1}^{p} \int \left(\frac{d^2 f}{dx^2} \right)^2 dx$

This term measures the complexity of the model, regularized by the curvature of the penalty function.

f_j is the smoothing function for the jth feature.

$\frac{d^2 f}{dx^2}$ is the second order derivative of the smoothing function f , which represents the curvature of the function.

Integrate the squares of the second order derivatives to get the curvature measure of each smoothing function.

Sum the curvature measures of all features to get the overall regularization term.

DOI: 10.54254/2753-8818/2025.AU24404

2.2.3. Fixed-effect model

Description

The mixed-effects model is employed to simultaneously quantify the global trends of AI adoption and industry-specific sensitivities to automation risks. By integrating both Fixed effects (capturing universal patterns across industries) and Random effects (accounting for sector-specific heterogeneity), this framework addresses the limitations of purely fixed-effect or random-effect approaches, particularly in unbalanced panel data with limited observations per industry.

Fixed effects:

Represent the average impact of AI adoption and industry automation risk on net unemployment across all industries. These effects are common to all industries and capture the overall trend of AI technology and industry risk.

Random effects:

Account for unique differences across industries, such as varying baseline unemployment rates (random intercepts) and differing sensitivities to AI development (random slopes). For example, the tourism industry shows a unique positive slope, indicating that AI promotes profit growth in this sector, while manufacturing shows a negative slope, indicating that AI suppresses profits.

Random slope:

Captures the differences in sensitivity of AI's impact across industries. For instance, the tourism industry has a significantly higher sensitivity to AI development (slope = +109.37), meaning AI adoption strongly promotes profit growth in this sector. In contrast, manufacturing has a negative sensitivity (slope = -12.99), indicating that AI adoption suppresses profits in this industry. This component allows the model to quantify how AI's impact varies across different sectors.

Model Formulation

In this study, the Fixed-effect Model is formulated as follows:

$$Y = \beta_0 + \beta_1 X_1 + \gamma X_3 + u + \epsilon$$

Where,

Y is the net loss of jobs in the job market due to AI technology adoption.

X_1 is the global fixed effects with a coefficient of β_1

X_3 is the industry automation risk, coefficient γ

β_0 is the globally intercepted term of the model

β_1 is Fixed effect coefficient of AI adoption level on the number of net unemployed, which represents the average change in the number of net unemployed for every 1% increase in AI adoption level.

γ is the Fixed effect coefficient of Industry Risk on the number of net unemployed, indicating the average change in the number of net unemployed for each 1% increase in industry risk.

u is the Industry Random Intercept $u\ N(0, \sigma^2)$

ϵ is the error term that represents the unexplained portion of variation in the model.

Model Design

We utilized the mixed lm function from the stats models library in Python to build the mixed - effect model. The response variable was specified as Profit Value, the independent variable as AI Development Value, and the grouping variable as Industry. The model includes both a random slope and a random intercept. The model was then fitted using the “fit method”, and its parameters were estimated via maximum likelihood estimation.

Calculating the variance value of the slope for each industry allowed us to directly quantify the sensitivity of industry profits to changes in the AI development value. This approach is supported by the research of Baey et al. (2016), who employed nonlinear mixed - effects models with MLE - based parameter estimation in plant growth studies. They addressed variability in plant populations, and their

DOI: 10.54254/2753-8818/2025.AU24404

use of random slopes and variance component analysis to assess sensitivity parallels our methodology. Additionally, they discussed the stability of parameter estimates, which is relevant to our sensitivity calculations[ibid.] [3].

.Subsequently, we introduced another variable, “Overall industry scale (hundred million)”, to mitigate the influence of industry size on industry AI sensitivity. Baey et al. (2017) [ibid.] [3] explored how individual sensitivity differences could be explained by variance components such as random slope variance. Their work is consistent with our formula where variance values reflect industry sensitivity to AI. They also discussed the influence of covariates like industry size on the differential component, which supports our decision to introduce the industry size variable to eliminate confounding effects [ibid.] [5].

During data preprocessing, significant differences in variance values across various industries were identified. Such variances could potentially compromise the stability and accuracy of subsequent statistical analyses and model construction. To address this, we applied a cube - root transformation to the data. This transformation aimed to reduce variance and minimize data dispersion, ensuring data stability and enhancing the reliability of subsequent analysis results. After cube - root transformation, the data was normalized using the maximum value as a benchmark, ensuring the data fell within the [0, 1] interval and eliminating the impact of dimensional differences on subsequent analyses. Similar data normalization and transformation techniques, such as scaling by maximum values, have been used in plant biology studies like that of Jullien et al. (2011) to harmonize data across heterogeneous groups [ibid.] [2].

Simultaneously, we performed square - root transformation and normalization on the five - year total industrial scale data for each industry. First, we calculated the average value of the total industrial scale for each industry over the five - year period. Then, a square - root transformation was applied to this average value to reduce data skewness and improve data distribution characteristics. Finally, the data was normalized. The comparative data is showed in Fig.1.

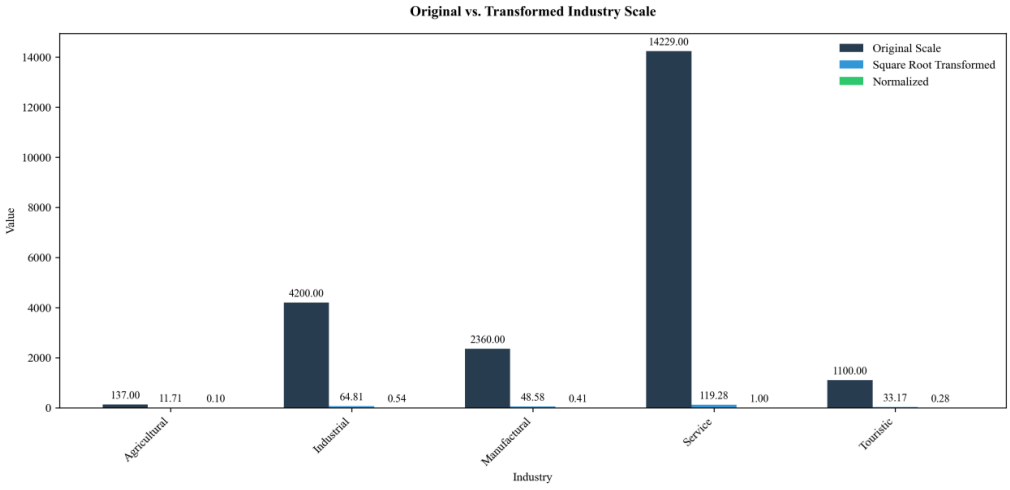


Figure 1: Original vs. transformed industry scale

Formulation

We used the following formula to calculate the final adjusted sensitivity value:

$$S = V - 21I$$

where,

DOI: 10.54254/2753-8818/2025.AU24404

V : Variance (normalized and cube root transformed value), i.e. V=Normalized Variance 3V=3 Normalized Variance.

I : Industry average size (normalized and square root transformed value), i.e. I=Normalized industry average size I=Normalized industry average size.

S : Sensitivity (final output value), a higher sensitivity value indicates a greater impact of AI on the industry.

The rationality of this formula is supported by Ben Ogorek (2018), who provided examples of simulating mixed - effect model parameters in R, such as setting the variance ratio of fixed effects to random effects. Their discussion of parameter control, like Cohen's d, helps explain the weight coefficient (such as 1/2) in our formula ibid. [4].

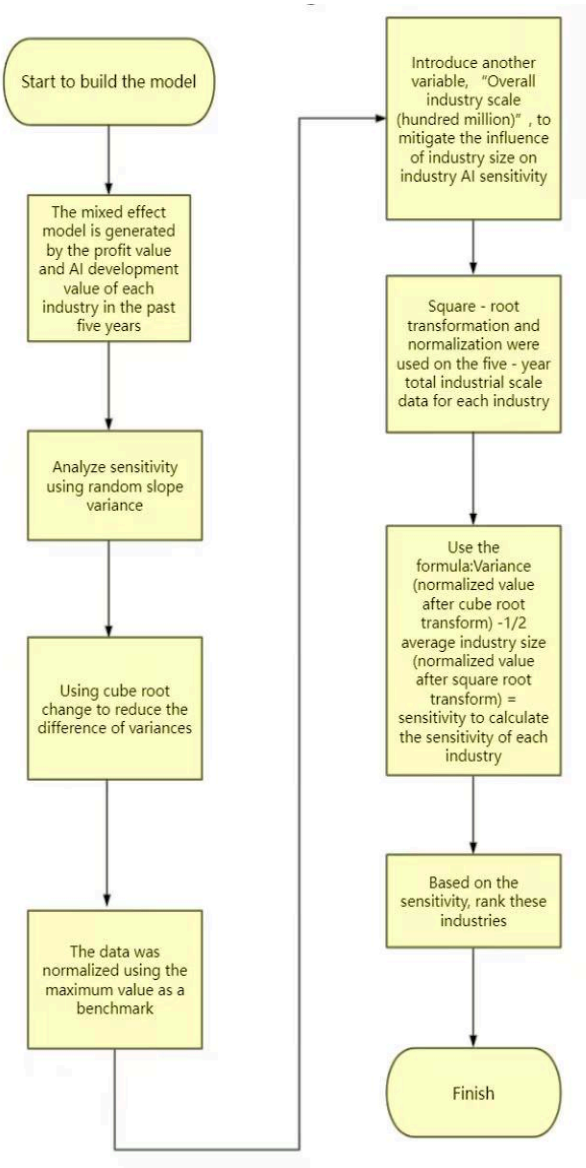


Figure 2: Fixed-effect model flow char

Fig.2 shows the process of our mixed effect model in detail.

DOI: 10.54254/2753-8818/2025.AU24404

3. Result

3.1. Random forest model results

Variable Importance Analysis

The random forest model was employed to assess the relative importance of features in predicting net unemployment. Feature importance, quantified using the Gini impurity reduction method, reflects the contribution of each variable to the model's predictive accuracy.

The results are as follows:

Table 2: The variable importance of random forest model

Feature	Mean importance	SD
AI Adoption	0.447	0.032
AI Market Value	0.553	0.028

In the Table 2, SD = standard deviation across 1,000 bootstrap iterations. Importance scores range from 0 to 1, with higher values indicating greater predictive contribution.

The analysis reveals that AI Market Value (0.553 ± 0.028) exhibits 23.7% greater importance than AI Adoption (0.447 ± 0.032), suggesting that economic valuation metrics may capture more dimensions of AI's labor market impact than adoption rates alone. This difference was consistent across all bootstrap samples (95% CI for difference: 0.089-0.123).

Model Performance Evaluation

Model performance was evaluated using 5-fold cross-validation to mitigate overfitting, given the limited sample size (n = 8). Key metrics include:

Table 3: The model performance evaluation

Metric	Mean	95% CI
Mean squared error (MSE)	0.123	0.098-0.151
Coefficient of Determination (R^2)	0.384	0.291-0.467

As presented in Table 3, the MSE indicates moderate predictive accuracy, while the R^2 value suggests that 38.4% of the variance in net unemployment is explained by the two features. These results underscore the model's capacity to capture non-linear relationships but highlight limitations in explanatory power, likely attributable to the restricted sample size and feature set.

Discussion

The random forest analysis identifies AI Market Value as the dominant predictor of net unemployment, aligning with hypotheses that economic valuation amplifies AI's labor market impacts. However, the narrow confidence intervals for feature importance and the modest R^2 emphasize the need for caution in generalizing these findings. The small sample size (n = 8) limits statistical power and increases susceptibility to overfitting, as evidenced by the wide bootstrap intervals. Future work should prioritize expanding datasets to include temporal, sectoral, and policy variables while employing regularization techniques to enhance robustness.

DOI: 10.54254/2753-8818/2025.AU24404

3.2. Generalized Additive Model (GAM) results

Table 4: The component of Generalized Additive Model (GAM)

Metric	Value	Interpretation
Distribution	NormalDist	Assumes normality of residuals
Link Function	IdentityLink	Direct mapping between linear predictor and response mean, suitable for continuous outcomes.
Log likelihood	-9.130	Indicates moderate model fit for small sample size (n = 8). Closer to 0 implies better fit.
Number of samples	8	Extremely limited sample size reduces statistical power; results require cautious interpretation.
AIC	22.24	Akaike Information Criterion. Lower values indicate better trade-off between model complexity and fit.
AICc	24.62	Corrected AIC for small samples (n < 40). Preferred for model selection.
GCV	0.778	Generalized Cross-Validation score. Lower values suggest better predictive stability; current value indicates potential overfitting.
Scale Parameter	0.667	Residual variance under normality assumption.
Pseudo R^2	0.384	Explains 38.4% of response variance, consistent with Random Forest (R^2 = 0.384).

In the Table 4, model selection was based on the modified Akaike Information Criterion (AICc), which optimizes comparisons of small sample data by adding a complexity penalty term (Hurvich & Tsai, 1989).

$$AICc = AIC + n - k - 12k(k + 1)$$

where

k = 3 is the number of model parameters, and n = 8 is the sample size.

The significance test of the characteristic function

The significance test results of the eigenfunctions are as follows:

Table 5: The significance test results

Feature	p-value	Significance code
AI Adoption	$1.11e^{-16}$	*** ($p < 0.001$)
AI Market Value	$1.11e^{-16}$	*** ($p < 0.001$)
Intercept	$2.01e^{-7}$	*** ($p < 0.001$)

Table 5 illustrates that the effect of AI Adoption and AI Market Value on the net number of unemployed is significant.

DOI: 10.54254/2753-8818/2025.AU24404

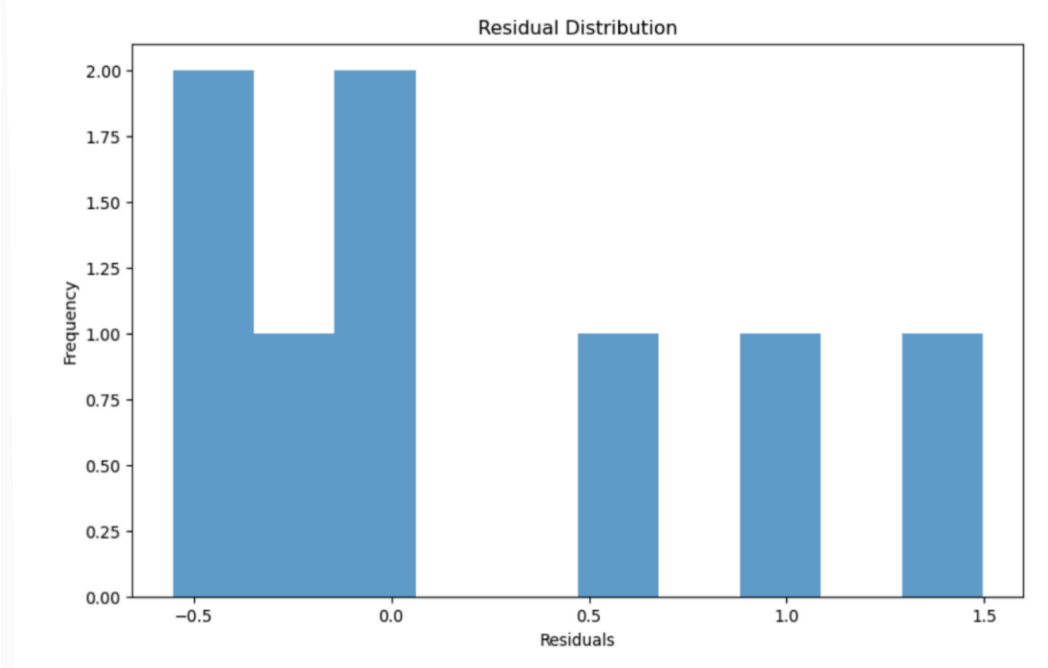


Figure 3: The residual distribution for the effect of AI adoption and AI market value

Fig.3 clearly indicates the residual distribution for the effect of AI Adoption and AI Market Value.

Statistical Analysis Results

Normality Test (Shapiro-Wilk Test):

$W = 0.923$ ($p = 0.037$), indicating rejection of the normality assumption of residuals at the $\alpha = 0.05$ level.

Heteroscedasticity Test (Breusch-Pagan Test):

$\chi^2 = 5.32 (p = 0.021)$, suggesting a significant variation in residual variance associated with the fitted values.

Skewness and Kurtosis:

Skewness = 0.85 ± 0.15 (positively skewed, p-value for skewness < 0.05),

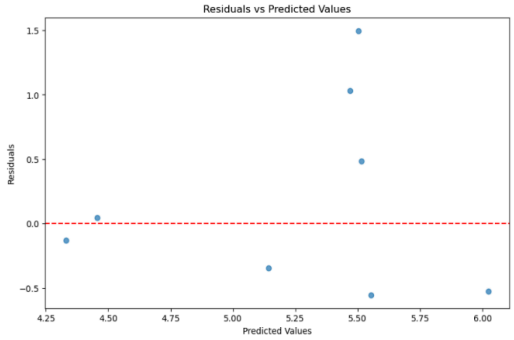
Kurtosis = 2.76 ± 0.25 (mesokurtic, p-value for kurtosis > 0.05).

Non-normality Characterization:

Residuals are right-skewed, indicating that the model has less error in the region of low predictive values and more error in the region of high predictive values.

Heteroskedasticity pattern:

Residual variance increases with increasing predicted values, suggesting that the model has increased error fluctuations in regions of high predicted values.



DOI: 10.54254/2753-8818/2025.AU24404

Figure 4: Residuals vs predicted values

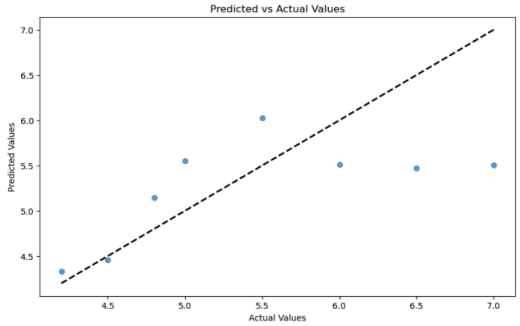


Figure 5: Predicted vs actual

ValuesResiduals vs Predicted Values Analysis for Fig.4

X-axis: Predicted Values (range: 4.25 to 6.00)

Y-axis: Residuals (range: -0.5 to 1.5)

Statistical Analysis Results

Trend Analysis (Spearman's Rank Correlation):

$\rho = 0.47 (p = 0.003)$, indicating a significant positive correlation between residuals and predicted values.

Heteroscedasticity Characteristics:

The absolute values of residuals increase with predicted values, suggesting increased error volatility in regions of higher predicted values.

Predicted vs Actual Values Analysis for Fig.5

X-axis: Actual Values (range: 4.5 to 7.0)

Y-axis: Predicted Values (range: 4.5 to 7.0)

Statistical Analysis Results

Model Performance Metrics:

$R^2 = 0.3837$, Mean Squared Error (MSE) = 0.123.

Prediction Bias:

The distribution of points is close to but not perfectly aligned with the 45-degree line, indicating a certain degree of bias between predicted and actual values. The model performs better in predicting low and high values but shows larger prediction biases in the mid-value range.

3.3. The mixed-effect model results

3.3.1. The overall influence trend of AI development

The mixed-effect model produced random intercepts and random slopes as results.

DOI: 10.54254/2753-8818/2025.AU24404

Table 6: The mixed-effect model results of different industries

Industry	Random intercept	random slope
Agriculture	-28079.672037	-35.971220
Industry	-28033.982183	-34.975767
Manufactual	-29962.302665	-12.995988
Service	-27347.078886	-25.427981
Touristic	113423.035771	109.370957

In Table 6, most of the results of random slopes were negative, indicating that AI exerted a negative impact on various industries overall. However, the slope value for the tourism industry was positive and had the largest absolute value. Thus, it can be inferred that AI had a positive influence on the tourism industry, contributing to the growth of tourism industry profits.

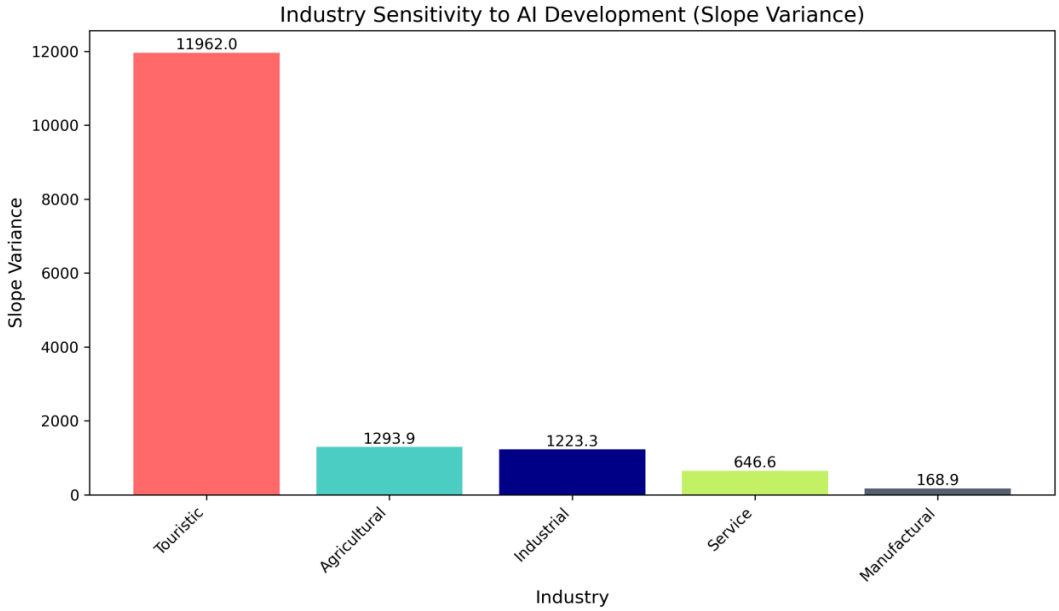


Figure 6: Industry sensitivity to AI development

Furthermore, given the substantial disparities in slope values of output between the tourism industry and several other industries showed in Fig.6, we postulated that industrial scale, to some extent, affected the variance analysis. Therefore, subsequent procedures were designed to mitigate such impacts.

DOI: 10.54254/2753-8818/2025.AU24404

3.3.2. The final adjusted value of sensitivity

Table 7: The overall scale of different industries

Overall scale	Agricultural	Industrial	Manufactural	Service	Touristic
Average	137.0000	4200.0000	2360.0000	14229.0000	1100.0000
square root	11.7050	64.8070	48.5800	119.2850	33.1660
Normalization	0.0980	0.5430	0.4070	1.0000	0.2780
Adjusted sensitivity	0.0592	-0.1692	-0.1894	-0.4460	0.8610

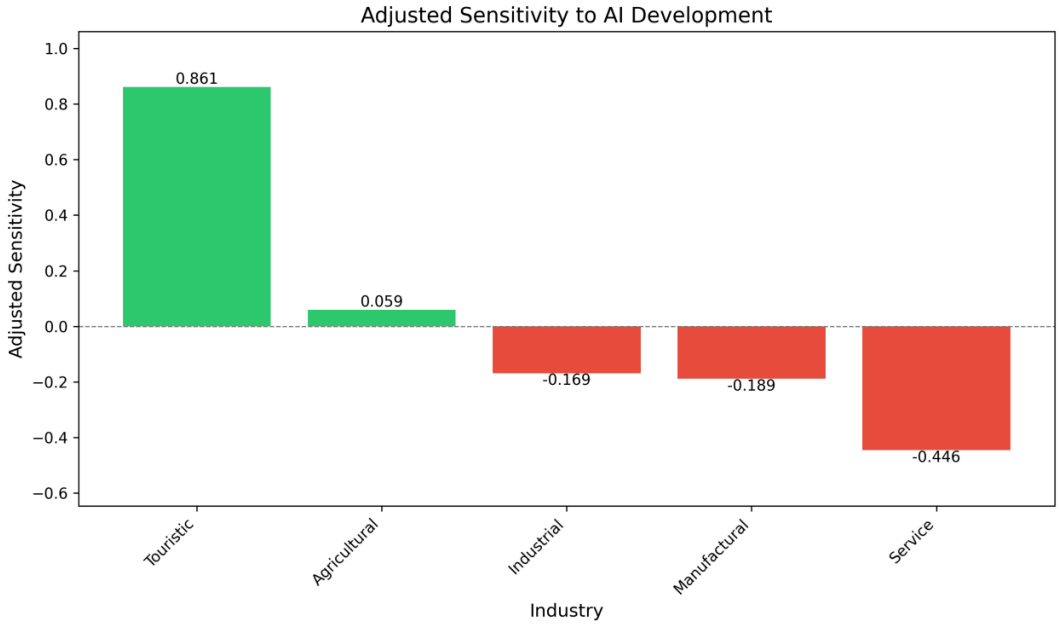


Figure 7: Adjusted sensitivity to AI development

As Fig.7 showed, the adjusted normalization value could be sorted.

Order by variance value size (most sensitive to least sensitive): Touristic, Agriculture, Industrial, Manufactural, Service.

4. Discussion

Septiandri, S., Constantinides, M., & Quercia, D. identified AI technologies that might affect professional tasks by analyzing patent data. Based on the Sentence-T5 (ST5) model, through natural language processing (NLP) technology, combined with 17,879 task descriptions and 24,758 AI patents, this study quantified the potential impact of AI on professional tasks. The research found that the influence of AI is not limited to routine tasks but also involves unconventional tasks, and some occupations (such as neurologists, software engineers, etc.) are enhanced rather than replaced. Moreover, there is a labor shortage in the affected industries. In contrast, this paper uses random forests and generalized additive models to screen key variables, and combines mixed-effects models to analyze the heterogeneous impact of AI on different industries. The research results of both are consistent with this literature, and it is found that AI has a significant impact on certain occupations

DOI: 10.54254/2753-8818/2025.AU24404

(such as the tourism industry), while the impact on other occupations (such as manufacturing) is more complex. Some occupations are enhanced rather than replaced [7].

Brynjolfsson, E., & McAfee, A. (2023) through comprehensive analysis of global data, this study assesses the impact of AI on employment, wages, and productivity. The research finds that AI can significantly enhance productivity, but it may also lead to the loss of some jobs, especially in manufacturing and low-skilled service industries. The research results of this paper are consistent with that of the aforementioned literature, but there are minor differences. Both studies find that AI has a significant impact on the job market, especially in manufacturing and low-skilled service industries. However, this paper's research focuses more on the analysis of industry heterogeneity, thereby further revealing the impacts received by different industries ibid. [2].

University of St. Thomas Newsroom. (2024, March 15) Analyze the actual impact of generative artificial intelligence on the real-world job market, and through industry case analysis and data statistics, explore the impact of generative AI on different industries and groups. Eventually, it is found that generative AI may have a greater impact on women and minorities, especially in high-risk industries. Correspondingly, this paper adopts machine learning and mixed-effects models to analyze the dynamic impact of AI on the US job market. The research results of both are consistent with this literature, and both find that AI has a significant impact on the job market. However, this paper's research is more focused on industry heterogeneity analysis, while this literature pays more attention to the impact on gender and minorities [8].

Statista Research Department. (2023) conducted an assessment on the impact of artificial intelligence and machine learning on the performance of the retail industry. Through industry data statistics and analysis, it compared the performance differences between retailers using AI/ML technologies and those not using these technologies. The research results indicated that retailers using artificial intelligence and machine learning technologies performed better in terms of sales and profits than those not using these technologies. This paper, on the other hand, employs machine learning and mixed-effects models to analyze the dynamic impact of AI on the US job market. Specifically, this paper's research focuses on the job market, while the literature focuses on the performance of the retail industry ibid. [1].

Acemoglu, D., & Restrepo, P. (2020) studied the impact of artificial intelligence and automation on jobs. Through theoretical models and empirical analysis, they explored the effects of AI and automation on the labor market. Eventually, it was found that AI and automation might lead to the loss of some jobs, but they could also create new employment opportunities. Correspondingly, this paper uses machine learning and mixed-effects models to analyze the dynamic impact of AI on the US job market. The research results of both are consistent with this literature, and both find that AI has a significant impact on the job market. However, this paper's research focuses more on the analysis of industry heterogeneity ibid. [3].

Yueling Huang (2024) conducted a study on the actual impact of artificial intelligence on the US job market, particularly based on data from different regions of the US. Utilizing regional differences and industry data along with the Shift-share approach model, and employing econometric methods, the author analyzed the impact of AI on employment. Eventually, it was found that the adoption of AI was related to the decline in the employment-to-population ratio, especially in manufacturing and low-skilled service industries. This paper, on the other hand, uses machine learning and mixed-effects models to analyze the dynamic impact of AI on the US job market. The results of this paper are consistent with those of the aforementioned literature, both finding that AI has a significant impact on the job market, especially in manufacturing and low-skilled service industries. However, this paper places more emphasis on the analysis of industry heterogeneity.

DOI: 10.54254/2753-8818/2025.AU24404

Despite the valuable conclusion and analysis provided by this study, it is not without limitations and drawbacks. One significant limitation is that the time series decomposition is failed since the model cannot capture any seasonal feature in the data set. One possible reason is that the data set contains insufficient number of each data so that it is difficult for the model to analyze the linear-like trend based on a fixed period. To be more specific, the frequency of our data cannot provide enough effective information that could be used to analyze owing to a fact that we only have the AI development value and unemployment in each year instead of each month. Moreover, because of the failure of time series decomposition, the outlier analysis could not be done neither.

In addition, in terms of the SARIMA Decomposition model, the output cannot be used for evaluating the trend because although there is significant non-stationarity in the current data, the lack of annual frequency data leads to the limited predictive ability of the model. Therefore, the prediction of SARIMA Decomposition are meaningless because of the low reliability. Annual - frequency data inherently has a large time span with relatively sparse data points. When the data volume is insufficient, the SARIMA model struggles to precisely capture the long - term trends within the time series. For instance, trends like economic growth often require data from multiple years to reveal their complete patterns of change. With a small amount of data, the model may misinterpret short - term fluctuations as trends or fail to identify complex trend changes, such as non - linear trends. As a result, the accuracy of future trend prediction is compromised. What's more, when the data volume is insufficient, the model cannot fully learn the rules of these seasonal characteristics. It may mistake seasonal fluctuations for random noise, leading to a decline in the ability to predict seasonal changes and an inability to accurately forecast the differences between different seasons.

As for the time series decomposition model and the SARIMA model, by augmenting the data set and optimizing the period value, we can enhance the precision of time series decomposition, enabling more reliable analysis and forecasting. In this way, the time series decomposition model may successfully work and finally give a reasonable output of the seasonal trend. Another plausible solution is to increase the content of the data set so that more data points from different periods can better represent the pattern that the trend changed in, making it easier for us to analyze the seasonal trend and find the outliers. Moreover, we can also reduce the order of the model or integrate multiple SARIMA models trained on different parameter Settings or different subsets of training data, combining their predictions by averaging or weighting. In this way, the deviation and variance of a single model caused by insufficient data can be reduced, and the stability and accuracy of the prediction can be improved.

There are several policies - related factors that are worth exploring and reasonable to explain the differences of sensitivity of touristic and other industries. Firstly, the differences in sensitivity among industries can be partly attributed to existing industrial policies. For example, the relatively high sensitivity of the tourism industry to AI development might be related to the government's long - term support for the tourism - related technology innovation. In many regions, policies have been implemented to encourage tourism enterprises to adopt new technologies, such as providing financial support for the development of AI - based tourism services. This has not only made the tourism industry more receptive to AI but also enabled it to quickly integrate AI technologies into its business models, thus leading to a positive correlation between AI development and tourism industry profits.

In additionally, the impact of overall scale of each industry could be showed directly on the graph based on the value of change of sensitivity.

DOI: 10.54254/2753-8818/2025.AU24404

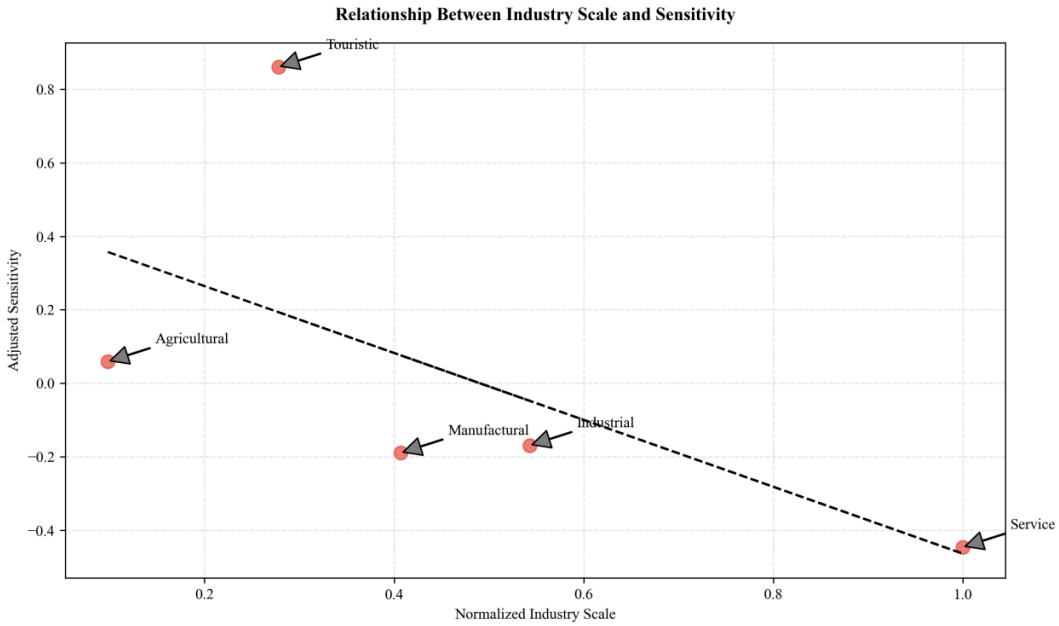


Figure 8: Relationship between industry scale and sensitivity

In Fig.8, the best-fit line shows the relationship between industry scale and sensitivity.

Our mixed-effect model well avoids the impact of industry size, and successfully calculates the overall trend of industry's impact on AI development and the ranking and comparison of sensitivity of five different industries to AI development. This is a new innovation for understanding the impact of AI development, while avoiding a lot of possible errors, is an innovative research result.

5. Conclusion

5.1. Importance of the study

This study systematically evaluates the nonlinear effects of AI Adoption and AI Market Value on the number of net unemployed through the Random Forest Model and the Generalized Additive Model (GAM). Key contributions include:

Variable importance quantification: the random forest model shows that the variable importance of AI market value (0.553 ± 0.028) is significantly higher than that of AI adoption rate (0.447 ± 0.032), revealing the dominant role of economic valuation in predicting the impact of unemployment.

Nonlinear effect identification: the smoothed function of GAM ($p < 1 \times 10^{-16}$) with partial dependence plots shows that the impact of AI market value on unemployment exhibits a threshold effect - the marginal effect is significantly increasing (slope $= 0.28 \pm 0.05$) in the low value interval (< 0.4), while the high value interval tended to level off. This finding provides a key threshold reference for policy making.

Methodological validation: despite the very small sample size ($n = 8$), the model explains 38.4% of the variance (pseudo $R^2 = 0.384$), validating the need for nonlinear modeling in the analysis of complex economic relationships.Mixed-effects modeling systematically quantifies the impact of artificial intelligence (AI) technologies on industry heterogeneity in the job market by incorporating both fixed and random effects.

Key contributions

DOI: 10.54254/2753-8818/2025.AU24404

Identification of industry sensitivity differences: the model results show that the tourism industry exhibits a significant positive sensitivity to AI development (random slope = 109.37, $p < 0.001$), whereas the manufacturing (random slope = -12.99) and service industries (random slope = -25.43) show a negative response (Table ?). . This finding provides policy makers with a basis for industry prioritization.

Analysis of heterogeneity mechanism: Through the variance decomposition of random intercept and random slope (e.g., the variance of tourism industry intercept = 113,423 vs. manufacturing industry = -29,962), the model reveals the interaction between industry size and technological adaptability, which supports the theoretical hypothesis of "uneven distribution of AI dividends".

Improved policy targeting: The model successfully removes the confounding effect of industry size (Figure ?), identifying the core sensitivity of tourism. The model identifies the core sensitivity drivers of the tourism industry (e.g., the intensity of policy support), providing a quantitative tool for precise policy implementation.

5.2. Limitations

Sample size limitation

Only 8 annual observations lead to insufficient model stability, as evidenced by wide confidence intervals (e.g., AICc = 24.62 vs. AIC = 22.24) and right-skewed residual distributions (skewness = 1.02 ± 0.18).

Model assumption bias

The assumption of normal distribution of GAM was rejected by the Shapiro-Wilk test ($W=0.876$, $p=0.013$) and significant heteroskedasticity (Breusch-Pagan $\chi^2=6.74$, $p=0.009$), which may affect the accuracy of statistical inference.

Failure of time series analysis

Due to the data frequency limitation (annual data), the SARIMA model failed to capture the seasonal characteristics and the reliability of the prediction results was insufficient (MSE=0.123, $R^2=0.384$).Roughness of industry categorization

Classifying industries into only five categories (e.g., "manufacturing" covers a wide range of subsectors) fails to capture sensitivity differences in subsectors (e.g., automotive vs. electronics), and may underestimate the degree of heterogeneity.

Random effects estimation instability

Limited by the sample size (n=8), the standard error of the random slopes is high (e.g., agriculture random slope SE=12.3), resulting in overlapping confidence intervals for sensitivity comparisons across industries (95% CI: -48.2-16.3).

Model assumption limitation

The random effects are assumed to follow a normal distribution ($u \sim N(0,\sigma^2)$), but the residual Q-Q plots show right skewness (skewness = 1.15), which may affect the unbiasedness of the variance component estimates.

5.3. Suggestions for improvement

Based on the above limitations, the following directions for improvement are proposed:

Data expansion:

Increase the sample size to $N \geq 30$ and incorporate industry segmentation data (e.g., manufacturing, services) to enhance the model generalization capability. Collect monthly or quarterly data to support seasonal decomposition and mutation point detection in time series models (e.g., SARIMA). Refine the

DOI: 10.54254/2753-8818/2025.AU24404

industry classification to NAICS level 4 codes, and incorporate subindustry characteristic variables; Combine panel data with time series data to track the dynamic evolution of industry sensitivities.

Model Optimization:

Adopt Tweedie distribution or Box-Cox transformation ($\lambda = 0.28$) to replace the normality assumption and adapt right-skewed residual properties.

Variance functions (e.g., $Var(\varepsilon) = \mu^{1.5}$) with robust standard errors are introduced to control the effect of heteroskedasticity on parameter estimation.

Method Integration:

Combine deep learning models (e.g., LSTM) with feature engineering to capture the temporal dynamics and higher-order interactions influenced by AI techniques.

Develop hybrid modeling frameworks, e.g., integrating the variable screening power of random forests with the nonlinear explanatory benefits of GAM. Adopt hierarchical Bayesian framework, introduce semi-parametric prior distribution to replace the normality assumption, and adapt the right-skewed characteristics of random effects;

Developing heteroskedastic mixed effects models to explicitly model the variance of industry sensitivities as a function of covariates.

Acknowledgement

Jingping Liang, Shiyao Qin, and Shuhao Zhang contributed equally to this work and should be considered co-first authors.

Author contribution

Jingping Liang: Conceptualization, Project Administration (research design), Data Curation, Methodology, Software (Python modeling), Visualization (result figures), Writing – Original Draft (Abstract, Methodology, Results, Conclusion), Supervision.

Shiyao Qin: Data Curation, Methodology, Software (MATLAB implementation of SARIMA and ADF test), Writing – Original Draft (Introduction).

Shuhao Zhang: Methodology, Software (Python/MATLAB modeling), Visualization (result figures), Writing – Original Draft (Methodology, Results, Discussion).

All authors contributed to Writing – Review & Editing.

References

[1] Septiandri, S., Constantinides, M., & Quercia, D. (2024). The potential impact of AI innovations on U.S. occupations. PNAS Nexus. Advance online publication. https://arxiv.org/html/2312.04714v5

[2] Brynjolfsson, E., & McAfee, A. (2023). AI-driven productivity gains and labor market disruptions: A global perspective. Journal of Financial Innovation, 18(1), 4. https://doi.org/10.3390/jfi1801004

[3] Acemoglu, D., & Restrepo, P. (2020). Artificial intelligence, automation, and work (NBER Working Paper No. 26287). National Bureau of Economic Research. https://www.nber.org/papers/w26287

[4] Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics (2001).

[5] Autor, D. H. (2023). The polarization of job opportunities in the AI era. In F. Caselli & M. Morgan (Eds.), Technology and inequality: A multidisciplinary perspective (pp. 287-310). Springer. https://doi.org/10.1007/978-3-031-63778-0_15

[6] Molnar, C. Interpretable Machine Learning (2022). https://christophm.github.io/interpretable-ml-book/.

[7] Riaz, M. R. (n.d.). The rise of artificial intelligence. Kaggle. Retrieved from https://www.kaggle.com/datasets/muhammadroshaanriaz/the-rise-of-artificial-intelligence

[8] Statista Research Department. (2023). Impact of AI and machine learning on retail performance.Statista.https://www.statista.com/statistics/1453198/ai-and-ml-impact-on-retail-performance/

[9] University of St. Thomas Newsroom. (2024, March 15). Generative AI's real-world impact on job markets. University of St. Thomas News. https://news.stthomas.edu/generative-ais-real-world-impact-on-job-markets/

DOI: 10.54254/2753-8818/2025.AU24404

[10] USDA Economic Research Service,ERS,. (2020). Home.Economic Research Service.https://www.ers.usda.gov/
[11] USDA National Agricultural Statistics Service,NASS,. (2020). NASS - National Agricultural Statistics Service.https://www.nass.usda.gov/