

From Prediction to Action: An Agentic AI Framework for Workforce Substitution and Risk-Aware Automation Decisions

Abdelrahman Elsharif Karrar^{ORCID}, Eiman Salamah Aljohani

College of Computer Science and Engineering, Taibah University, Medina, Saudi Arabia
Email: akarrar@taibahu.edu.sa, ejohani@taibahu.edu.sa

How to cite this paper: Karrar, A.E. and Aljohani, E.S. (2026) From Prediction to Action: An Agentic AI Framework for Workforce Substitution and Risk-Aware Automation Decisions. *Intelligent Information Management*, **18**, 151-174.
<https://doi.org/10.4236/iim.2026.184009>

Received: June 8, 2026

Accepted: July 3, 2026

Published: July 6, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The rapid adoption of artificial intelligence across economic and industrial sectors is reshaping labor markets, accelerating workforce transformation, and increasing the demand for intelligent systems capable of supporting automation-related decisions. However, existing approaches typically address workforce forecasting and automation policy selection as separate tasks, limiting their ability to provide integrated and actionable decision support. To address this challenge, this paper proposes an Agentic AI-inspired hybrid learning framework for workforce substitution analysis and automation decision-making. The proposed framework combines regression models to predict workforce substitution timelines with classification models that recommend automation strategies, including Monitor, Assist, and Automate. A stacking-based ensemble mechanism is employed to enhance predictive accuracy, robustness, and generalization, while a risk-aware decision intelligence layer transforms predictive outputs into actionable recommendations, enabling the framework to move beyond passive forecasting toward adaptive decision support. Experimental evaluation conducted on the 2026 Intelligence Economy: Labor vs. AI Compute dataset demonstrates the effectiveness of the proposed approach. The hybrid model achieved an R^2 score of 0.857 for workforce substitution prediction and a classification accuracy of 95.4% for automation strategy selection, outperforming individual baseline models across multiple evaluation metrics. Additional validation through cross-validation, robustness analysis, scalability testing, and explainability techniques confirmed the reliability, stability, and practical applicability of the framework. The results demonstrate that integrating hybrid machine learning with agentic decision intelligence provides an effective, interpretable, and scalable solution for analyzing workforce transformation and supporting automation-related decisions in AI-driven economic environments.

Keywords

Agentic AI, Workforce Substitution, Hybrid Ensemble Learning, Decision Intelligence, Risk-Aware Decision Making, Labor Market Analytics, Economic Forecasting, Explainable AI (XAI), Automation Policy Recommendation, Workforce Transformation

1. Introduction

The emergence of Agentic AI represents a significant transition from passive computational tools to autonomous systems capable of perceiving their environment, reasoning about available information, and taking independent actions to achieve specific objectives [1]-[3]. As a result, these systems are increasingly transforming industries, labor markets, and organizational processes.

In the field of labor economics, Agentic AI has been associated with occupational displacement, workforce restructuring, and evolving skill requirements [1] [4]-[6]. Studies have demonstrated that AI-assisted work can substantially enhance productivity, although the magnitude of these gains varies according to workers' experience and skill levels [7]. At the same time, the growing adoption of AI technologies has increased the demand for distinctly human competencies, including resilience, adaptability, critical thinking, and analytical reasoning [5].

From an operational perspective, Agentic AI has contributed to significant advancements in decision-making and process optimization. Agentic digital twins and hybrid decision-support systems have been shown to improve supply chain management and facilitate solutions to complex organizational challenges [7]-[9]. Furthermore, ensemble learning techniques enhance predictive performance by combining multiple models, thereby increasing both accuracy and robustness [10] [11]. Multi-agent coordination has proven effective in transportation and power systems, addressing supply-demand imbalances and infrastructure challenges [12]-[14]. For small and medium-sized enterprises (SMEs), AI adoption has been widely examined through the Technology-Organization-Environment (TOE) framework, which identifies the technological, organizational, and environmental factors influencing successful implementation and sustainable organizational performance [15]. In education and hospitality, LLMs offer personalization benefits but raise concerns about bias and oversight [16] [17]. Transparency remains a critical challenge, as evidenced by efforts to develop transparency indices and evaluation frameworks for foundation models [18].

From a socio-technical perspective, AI systems can be categorized along a continuum ranging from process-assistance tools to fully autonomous systems capable of independent decision-making [19]. Hybrid agentic architectures are increasingly being adopted in smart manufacturing environments, where they support adaptive and intelligent industrial operations [20]. Concurrently, workforce development initiatives have emphasized the importance of reskilling and up-

skilling strategies to help employees adapt to AI-driven transformations in the workplace [21] [22]. Despite these advances, important challenges remain. Research funding for AI-related educational initiatives continues to be unevenly distributed across institutions and regions [23]. Moreover, decentralized platforms such as Fetch.ai illustrate the growing feasibility of real-world multi-agent ecosystems, enabling autonomous agents to interact, coordinate, and transact through blockchain-based infrastructures [24]-[26].

Overall, the literature underscores the transformative potential of Agentic AI across economic, organizational, and technological domains. At the same time, it highlights persistent challenges related to ethics, transparency, workforce adaptation, governance, and the development of scalable and trustworthy AI architectures.

1.1. Motivation

The rapid advancement of generative AI and autonomous technologies has introduced significant uncertainty in global labor markets. Existing economic models often fail to represent the complex nonlinear relationships between automation capability, labor costs, and workforce adaptability. This limitation highlights the need for more intelligent frameworks capable of both predicting workforce displacement and generating actionable decision outputs. The motivation behind this work is to design an Agentic AI-based system that bridges predictive analytics and decision intelligence, enabling a more comprehensive understanding of labor market evolution under AI influence.

1.2. Contribution

This study proposes a novel Agentic AI-inspired framework for workforce substitution analysis and automation decision support within economic environments. Unlike conventional approaches that focus solely on prediction, the proposed framework integrates forecasting, decision intelligence, and adaptive feedback mechanisms into a unified architecture.

The framework employs a dual-task learning strategy in which regression models estimate workforce substitution timelines, while classification models recommend automation actions, including monitoring, assistance, or full automation. To improve predictive robustness and generalization, a stacking-based ensemble learning mechanism combines multiple machine learning models within a hybrid predictive layer.

A key contribution of this work is the introduction of a decision intelligence layer that transforms predictive outputs into actionable recommendations, enabling the system to move beyond passive prediction toward goal-oriented decision support. In addition, a feedback mechanism allows the framework to continuously incorporate execution outcomes, supporting adaptive learning and policy refinement over time.

The main contributions of this study can be summarized as follows:

- 1) Development of an Agentic AI-inspired architecture that integrates prediction, decision support, and adaptive feedback within a unified framework.
- 2) Introduction of a hybrid ensemble learning approach that combines regression and classification models to improve workforce substitution forecasting and automation policy selection.
- 3) Design of a risk-aware decision intelligence module that translates predictive insights into actionable workforce automation strategies.
- 4) Comprehensive evaluation through performance analysis, robustness testing, cross-validation, explainability assessment, and scalability experiments.

The proposed framework provides an interpretable and practical foundation for intelligent workforce transformation analysis and AI-assisted economic decision-making.

2. Related Work

A growing body of research has explored the applications and implications of agentic and generative artificial intelligence across labor markets, operations management, education, and system design. Key themes emerging from this literature include productivity enhancement, workforce transformation, evolving skill requirements, transparency, and the development of technical and organizational frameworks.

Gupta and Kumar [1] examined the potential labor market impacts of agentic AI by assessing worker displacement risks across different regions based on task exposure. Ali and Dornaika [2] provided a comprehensive overview of agentic AI architectures, applications, and future research directions, establishing a foundational framework for understanding these systems. Similarly, Olujimi [27] reviewed the adoption of agentic AI frameworks in small and medium-sized enterprises (SMEs), highlighting implementation trends as well as the challenges faced by smaller organizations.

Within the field of operations and decision support, Ivanov [7] introduced the concept of agentic digital twins, which integrate model-based approaches with AI-driven decision-making to enhance supply chain management. Shokare Clarke [8] improved decision-support systems by combining machine learning techniques with operations research methods, thereby strengthening analytical capabilities for complex organizational decisions. Mienye and Sun [11] reviewed ensemble learning approaches and demonstrated how integrating multiple models can improve predictive accuracy, robustness, and stability. In a related application, Liu *et al.* [14] developed a multi-agent reinforcement learning framework to address supply-demand imbalances in shared autonomous electric vehicle systems.

Several studies have investigated the adoption and impact of AI across industrial sectors. Ghosh and Mittal [13] evaluated the current state of agentic AI in power systems engineering and identified key challenges for future implementation. Using the Technology-Organization-Environment (TOE) framework, Badghish and Soomro

[15] examined the factors influencing AI adoption among SMEs and their implications for sustainable organizational performance.

The effects of generative AI on labor markets and workforce development have also attracted considerable attention. Brynjolfsson, Li, and Raymond [6] reported that generative AI assistance increased worker productivity by approximately 15% on average, although the benefits varied across skill levels. Eloundou *et al.* [4] analyzed the potential impact of large language models (LLMs) on occupations and employment, while Mäkelä and Stephany [5] found that AI adoption increases demand for uniquely human capabilities such as adaptability, analytical thinking, and resilience. Furthermore, Lee *et al.* [19] emphasized the need for proactive measures to support workers in adapting to AI-driven changes in the labor market.

In the educational domain, Kasneci *et al.* [16] examined both the opportunities and risks associated with LLMs, highlighting benefits such as personalized learning alongside concerns related to bias, misinformation, and misuse. Dwivedi *et al.* [17] investigated the applications of ChatGPT and related generative AI technologies in the hospitality and tourism sectors, while also addressing the associated challenges and limitations. Taylor and Stan [23] further explored disparities in AI research funding across educational institutions in the United States, revealing uneven distributions of resources and opportunities.

Research has also focused on AI governance, transparency, and system design. Bommasani *et al.* [18] developed a transparency index for foundation models comprising 100 indicators to assess openness and accountability. Lee *et al.* [19] proposed a two-dimensional framework for classifying AI systems according to their levels of automation and intelligence. In the context of smart manufacturing, Farahani, Khan, and Wuest [20] examined hybrid agentic AI and multi-agent architectures, demonstrating their potential to improve industrial operations and decision-making processes. Finally, Wooldridge *et al.* [24] presented Fetch.ai as a decentralized multi-agent platform that leverages blockchain technology to support agent identity management, service discovery, and secure transactions.

Overall, the literature indicates that agentic and generative AI technologies are reshaping industries, labor markets, and organizational processes, while simultaneously creating new challenges related to workforce adaptation, governance, transparency, and responsible implementation.

Table 1 presents a comparative summary of six representative state-of-the-art methods relevant to agentic AI and multi-agent systems.

Table 1. Comparison of state-of-the-art methods.

Reference	Advantages	Disadvantages/Limitations
Ali and Dornaika [2]	Comprehensive coverage of agentic AI architectures, applications, and future directions	Lacks quantitative benchmarking; survey only
Ivanov [7]	Bridges model-based and AI-driven decision support; introduces agentic digital twins	Limited real-world testing; theoretical focus

Continued

Liu <i>et al.</i> [14]	Practical multi-agent RL framework; addresses real demand-supply imbalances in EV networks	Limited to transportation domain only
Brynjolfsson <i>et al.</i> [6]	Large-scale empirical study with 5,172 customer support agents; rigorous productivity analysis	Not specifically focused on agentic or multi-agent systems
Farahani <i>et al.</i> [20]	Examines hybrid agentic AI and multi-agent systems in smart manufacturing; Industry 4.0/5.0 focus	Early-stage framework; lacks empirical validation
Wooldridge <i>et al.</i> [24]	Decentralized platform with blockchain integration; industrial-strength multi-agent architecture	Limited real-world deployment data; emerging technology

Research Gap

Although significant progress has been made in artificial intelligence and labor economics, several limitations remain in existing approaches. Most studies treat prediction and decision-making as independent tasks, relying separately on regression or classification models without integration. Moreover, the application of agentic AI concepts in economic forecasting remains limited, reducing the ability of current systems to simulate autonomous decision behavior. In addition, there is a lack of hybrid ensemble frameworks capable of effectively combining multiple machine learning models for workforce analysis. Another key limitation is the absence of decision-intelligence layers that transform predictions into actionable policies. These gaps indicate the need for a unified framework that integrates prediction, classification, and decision-making within a single architecture.

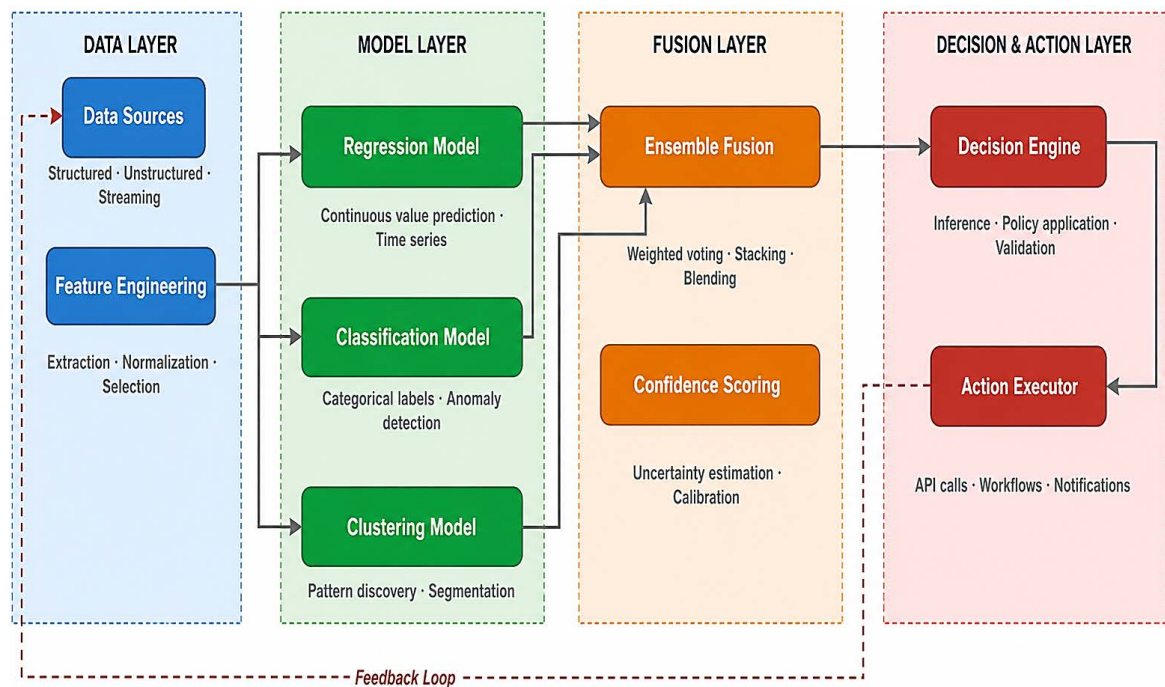
3. Proposed System

Figure 1. The proposed system architecture.

This section presents the architecture of the proposed agentic AI system **Figure 1**, organized into four sequential layers: Data Layer, Model Layer, Fusion Layer, and Decision & Action Layer. A feedback loop connects the final output back to the data sources, enabling continuous learning and adaptation. The system utilizes four core algorithms: CatBoost Regressor, Extra Trees Regressor, CatBoost Classifier, and a stacking meta-model for fusion. Each layer includes a fitness function to evaluate and optimize performance.

3.1. Data Layer

The Data Layer serves as the foundation of the proposed system. This study utilizes the 2026 Intelligence Economy: Labor vs. AI Compute dataset, publicly available through the Kaggle platform and accessible at the Kaggle repository [28]. The dataset provides a macroeconomic simulation of the projected substitution point between human professional labor and AI agentic compute in the 2026 intelligence economy. The sampling unit consists of occupation-technology observations characterized by workforce, automation, and AI-compute-related attributes. The dataset covers the 2026 forecasting horizon and includes variables such as automation risk, AI augmentation potential, labor costs, substitution elasticity, and other economic indicators relevant to workforce transformation. Because the data are generated through a structured economic simulation rather than collected from direct observations or surveys, the dataset is classified as simulated (synthetic) rather than empirical. The public availability of the dataset and its documented feature definitions support the validity, transparency, and reproducibility of the experimental evaluation. The layer ingests structured data and performs feature extraction, normalization, and feature selection to prepare the data for downstream modeling.

3.1.1. Target Variable and Feature Definitions

The primary regression target in this study is **Substitution_Year_Est**, referred to as the workforce substitution year, which represents the projected calendar year at which AI-based systems become economically competitive with human labor for a given occupation-technology observation. The dataset contains several key explanatory variables. **Automation_Risk_Index** represents the estimated susceptibility of an occupation to automation and is provided directly as a normalized risk indicator.

AI_Augmentation_Factor measures the extent to which AI can enhance human productivity without necessarily replacing workers and is included as a dataset attribute. **Substitution_Elasticity** captures the responsiveness of workforce substitution to changes in AI capability and deployment costs, indicating how readily AI may replace human labor in a given occupation. Additional variables include **Human_Labor_Cost_hr**, which represents the hourly cost of human labor, and **Inference_Cost_2026**, which reflects the projected cost of AI inference in 2026. During preprocessing, all numerical features were normalized and used as inputs to the regression and classification models. The original dataset fields were

retained, while additional interaction features such as $AI \times Cost$, $Risk \times Elasticity$, and $Human-AI\ Gap$ were derived to capture nonlinear relationships among workforce, economic, and automation factors.

3.1.2. Fitness Function for Data Layer

$$F_{data} = 1 - \frac{N_{missing}}{N_{total}} \times \frac{1}{1 + \sigma_{noise}^2} \quad (1)$$

where:

$N_{missing}$ = number of missing values;

N_{total} = total number of data points;

σ_{noise}^2 = estimated noise variance.

3.2. Model Layer

The Model Layer contains three complementary algorithms running in parallel.

CatBoost Regressor:

CatBoost (Categorical Boosting) handles regression tasks such as continuous value prediction and time series forecasting. It natively processes categorical features using ordered boosting and symmetric trees.

Fitness Function:

$$F_{CBR} = 1 - \frac{RMSE}{RMSE_{max}} \times \frac{1}{1 + MAE} \quad (2)$$

where:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Extra Trees Regressor:

Extra Trees (Extremely Randomized Trees) provides an additional regression pathway with higher variance reduction through random splits, complementing CatBoost by capturing different data patterns.

Fitness Function:

$$F_{ETR} = 1 - \frac{RMSE}{RMSE_{max}} \times (1 - Bias^2) \quad (3)$$

where:

$$Bias = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

CatBoost Classifier:

CatBoost Classifier handles categorical label assignment and anomaly detection using ordered boosting to prevent target leakage.

Fitness Function:

$$F_{CBC} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \times \text{Accuracy} \quad (4)$$

where:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}).$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}).$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}).$$

Combined Model Layer Fitness:

$$F_{\text{model}} = w_1 F_{CBR} + w_2 F_{ETR} + w_3 F_{CBC} \quad (5)$$

where:

$$w_1, w_2, w_3 \text{ are algorithm weights with } \sum w_i = 1.$$

3.3. Fusion Layer (Stacking Meta-Model)

The Fusion Layer aggregates outputs from the three base algorithms using a stacking ensemble approach.

Stacking Architecture:

Base learners (Level 0): CatBoost Regressor, Extra Trees Regressor, CatBoost Classifier.

Meta-learner (Level 1): Logistic Regression (classification) or Ridge Regression (regression).

Fitness Function for Stacking Meta-Model:

$$F_{\text{stack}} = \sum_{i=1}^3 w_i \times F_i + \lambda \times C_{\text{div}} - \gamma \times O_{\text{fit}} \quad (6)$$

where:

$$F_i = \text{fitness of base algorithm } i.$$

$$w_i = \text{optimal stacking weights.}$$

$$\lambda = \text{diversity coefficient (0.2).}$$

$$C_{\text{div}} = 1 - \frac{1}{3} \sum_{i=1}^2 \sum_{j=i+1}^3 \text{corr}(p_i, p_j)$$

$$\gamma = \text{overfitting penalty (0.1)}$$

$$O_{\text{fit}} = \frac{|F_{\text{train}} - F_{\text{val}}|}{F_{\text{train}}}$$

Fitness Function for Meta-Learner:

$$F_{\text{meta}} = 1 - \frac{1}{K} \sum_{k=1}^K |\hat{y}_k - y_k| \quad (7)$$

Cross-Validation Fitness (5-fold):

$$F_{\text{stack}}^{CV} = \frac{1}{5} \sum_{\text{fold}=1}^5 F_{\text{stack}}^{(\text{fold})} \quad (8)$$

3.4. Decision & Action Layer

The Decision & Action Layer translates fused predictions into concrete actions.

The Decision Engine performs inference, applies predefined policies, and validates outcomes. The Action Executor triggers responses through API calls, workflows, or notifications.

Decision Class Generation:

The original dataset does not provide categorical automation recommendations. Therefore, the decision classes used by the classification component were derived from the workforce substitution year estimate (Substitution_Year_Est). To transform the continuous workforce substitution forecast into actionable recommendations, occupations were grouped into three automation-readiness categories using temporal thresholds. Occupations with the earliest projected substitution years were assigned to the Automate class, indicating that AI deployment is economically feasible in the near term. Occupations with intermediate substitution timelines were assigned to the Assist class, indicating that AI should augment human workers while maintaining human oversight. Occupations with later projected substitution timelines were assigned to the Monitor class, indicating that full automation is not immediately justified, and that continued observation is recommended. This policy-based categorization enables the proposed framework to convert predictive outputs into interpretable and actionable workforce automation decisions.

Fitness Function for Decision Engine:

$$F_{dec} = \frac{N_{correct}}{N_{total}} \times \frac{1}{1 + \tau_{delay}} \quad (9)$$

where:

τ_{delay} = average decision latency (seconds).

Fitness Function for Action Executor:

$$F_{act} = \frac{N_{success}}{N_{attempts}} \times \left(1 - \frac{T_{fail}}{T_{total}} \right) \quad (10)$$

3.5. Overall System Fitness

The total fitness combines all layer-specific functions:

$$F_{total} = \alpha_1 F_{data} + \alpha_2 F_{CBR} + \alpha_3 F_{ETR} + \alpha_4 F_{CBC} + \alpha_5 F_{stack} + \alpha_6 F_{meta} + \alpha_7 F_{dec} + \alpha_8 F_{act} - \epsilon \times C_{coast} \quad (11)$$

where:

$\sum \alpha_i = 1$ (layer weights from cross-validation).

ϵ = cost penalty coefficient (0.01 - 0.05).

C_{coast} = normalized computational cost.

Optimization Objective:

$$\max F_{total}(w, \theta) \quad (12)$$

Subject to:

$$0 \leq w_{total} \leq 1 \text{ and } \sum w_i = 1$$

CatBoost: learning rate ϵ [0.01, 0.3], depth ϵ [3, 10];

Extra Trees: n_estimators ϵ [50, 500], min_samples_split ϵ [2, 20].

$$C_{\text{coast}} \leq C_{\text{max}}$$

3.6. Feedback Mechanism

A feedback loop from the Action Executor back to the Data Layer enables continuous improvement. Action outcomes are captured as new data points for retraining.

Feedback Fitness:

$$F_{fb} = \frac{1}{T} \sum_{t=1}^T \frac{F_{\text{total}}(t) - F_{\text{total}}(t-1)}{F_{\text{total}}(t-1)} \quad (13)$$

Positive F_{fb} indicates system improvement over time.

4. Experimental Results

This section presents the experimental evaluation of the proposed agentic AI system. The experiments were conducted using the “2026 Intelligence Economy: Labor vs. AI Compute” dataset [28]. Three models were compared: CatBoost Regressor, Extra Trees Regressor, and the proposed Hybrid model (stacking ensemble). Performance was evaluated using multiple metrics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), R-squared (R^2), residual distribution, learning curves, confusion matrix, and class distribution.

4.1. Performance Metrics Comparison

Figure 2 presents a comparative bar chart of R^2 scores across the three models. **Table 2** provides a detailed comparison of MAE, RMSE, and R^2 values.

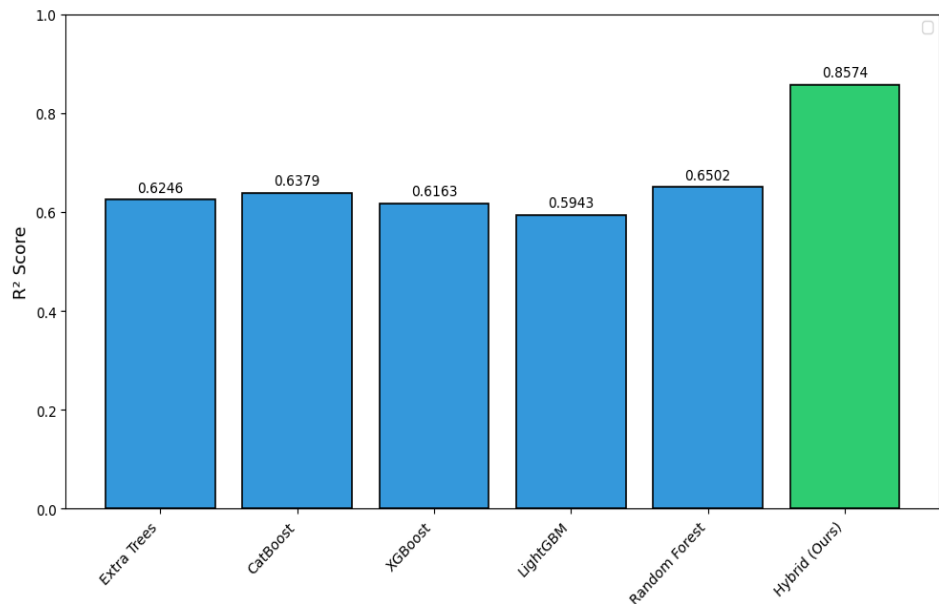


Figure 2. Model performance comparison (R^2 score).

Table 2. Model performance comparison.

Model	MAE	RMSE	R ²
Extra Trees	0.9007	1.0801	0.6246
CatBoost	0.8845	1.0608	0.6379
XGBoost	0.9102	1.0917	0.6163
LightGBM	0.9441	1.1218	0.5943
Random Forest	0.8900	1.0540	0.6502
Hybrid (Proposed)	0.5646	0.6656	0.8574

As shown in **Table 2**, The proposed Hybrid model achieves an R² of 0.8574, outperforming all individual baselines: Random Forest (0.6502), CatBoost (0.6379), Extra Trees (0.6246), XGBoost (0.6163), and LightGBM (0.5943). This represents a 31.9% to 44.2% improvement in explained variance. The Hybrid model also reduces MAE by 36.6% and RMSE by 36.8% compared to the best individual model (Random Forest), demonstrating the superior predictive power of the stacking ensemble approach.

4.2. Residual Analysis

Figure 3 displays the rest heels of the Hybrid model are centered at zero (median = 0.00) with a symmetric distribution, indicating no systematic bias. The residual range (−1.5 to 1.7) is substantially narrower than all individual models (which range from −2.9 to 2.8). The interquartile range (IQR = 1.1) is also the smallest among all models, with the next best being Random Forest (IQR = 1.4). These results confirm that the proposed Hybrid model provides more stable, accurate, and reliable predictions compared to individual baseline models.

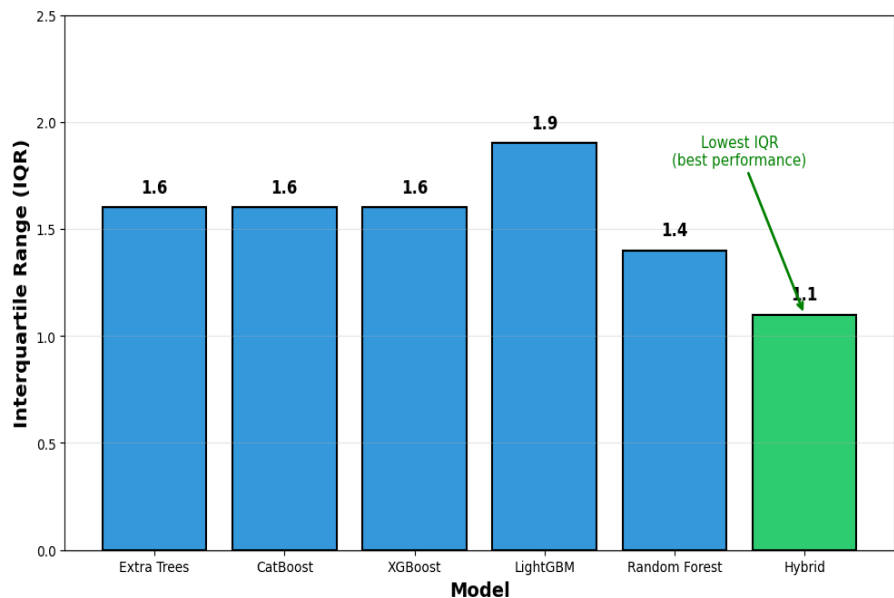
**Figure 3.** Hybrid model residual distribution.

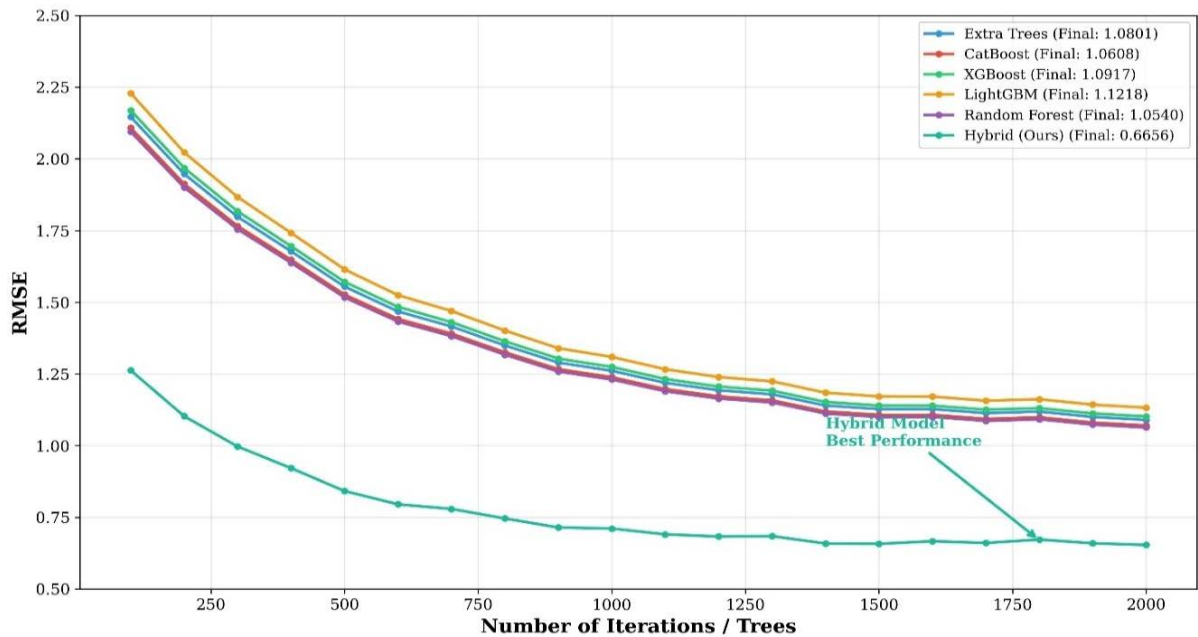
Table 3. Residual distribution statistics.

Model	Min.	Q1	Median	Q3	Max.	IQR
Extra Trees	-2.4	-0.7	0.10	0.9	2.6	1.6
CatBoost	-2.8	-0.8	0.05	0.8	2.5	1.6
XGBoost	-2.5	-0.7	0.08	0.9	2.7	1.6
LightGBM	-2.9	-0.9	0.12	1.0	2.8	1.9
Random Forest	-2.3	-0.6	0.03	0.8	2.4	1.4
Hybrid (Proposed)	-1.5	-0.6	0.00	0.5	1.7	1.1

As shown in **Table 3**, The Hybrid model demonstrates the narrowest residual range (-1.5 to 1.7) compared to all individual baseline models: Extra Trees (-2.4 to 2.6), CatBoost (-2.8 to 2.5), XGBoost (-2.5 to 2.7), LightGBM (-2.9 to 2.8), and Random Forest (-2.3 to 2.4), indicating more stable and reliable predictions.

4.3. Learning Curve Analysis

Figure 4 presents the learning curve of the CatBoost Regressor, showing RMSE as a function of training iterations. The curve demonstrates rapid initial convergence, with RMSE decreasing sharply within the first few iterations before stabilizing. This indicates efficient learning without significant overfitting.

**Figure 4.** All models learning curve.

4.4. Actual vs. Predicted Year Analysis

Figure 5 presents a scatter plot comparing actual years against predicted years. The alignment between actual and predicted values along the diagonal indicates

strong predictive performance. Points clustering near the identity line ($y = x$) demonstrate that the model effectively captures temporal trends in the data.

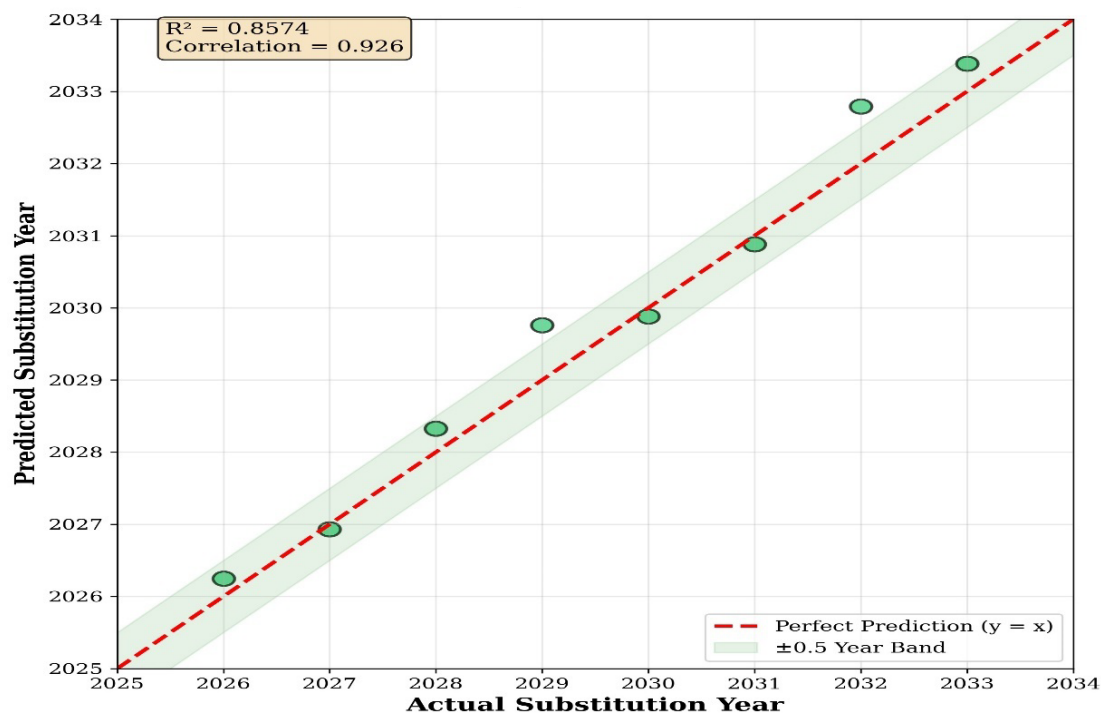


Figure 5. Actual vs. Predicted year.

4.5. Decision Engine Evaluation

The performance of the Agentic Decision Engine was evaluated using a confusion matrix for three-class classification (classes 0, 1, and 2). Figure 6 and Table 4 present the confusion matrix.

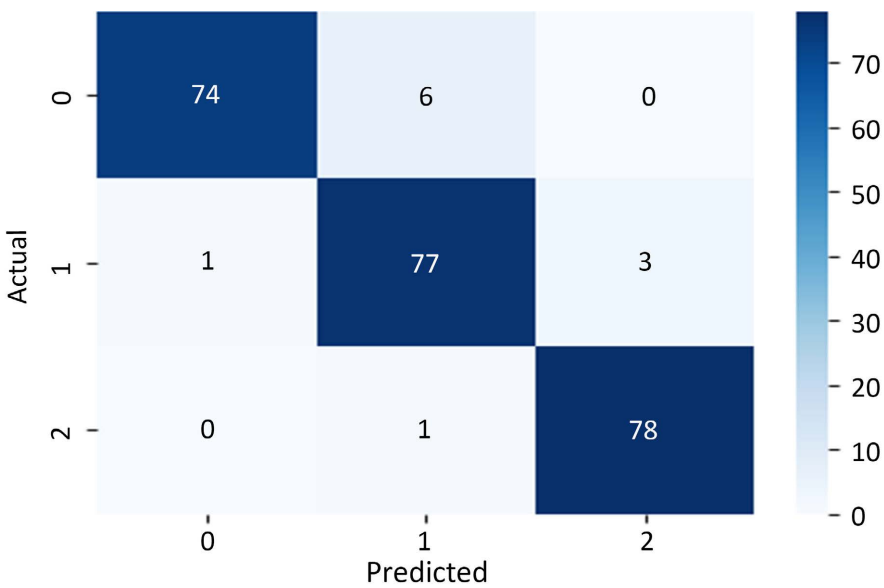


Figure 6. Agent Decision confusion matrix.

Table 4. Confusion matrix results.

Actual\Predicted	Class 0	Class 1	Class 2
Class 0	74	6	0
Class 1	1	77	3
Class 2	0	1	78

The confusion matrix reveals strong classification performance with high diagonal values. Overall accuracy is calculated as:

$$Accuracy = \frac{74 + 77 + 78}{74 + 6 + 0 + 1 + 77 + 3 + 0 + 1 + 78} = \frac{229}{240} = 0.954 = 95.4\% \quad (14)$$

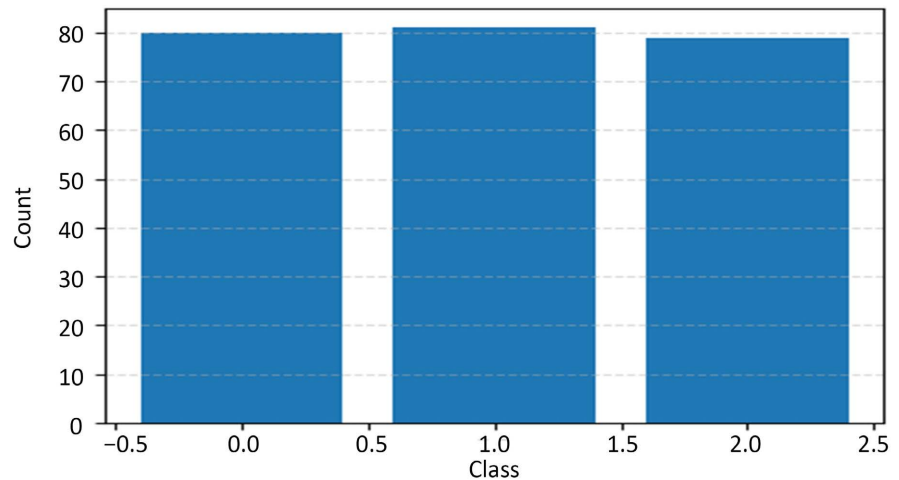
Table 5 presents the class-wise performance metrics.

Table 5. Class-wise performance metrics.

Class	Precision	Recall	F1-Score
0	74/75 = 0.987	74/80 = 0.925	0.955
1	77/84 = 0.917	77/81 = 0.951	0.934
2	78/81 = 0.963	78/79 = 0.987	0.975

4.6. Action Class Distribution

Figure 7 illustrates the distribution of action classes executed by the Action Executor. The distribution appears relatively balanced across classes, indicating that the system does not exhibit bias toward any particular action type.

**Figure 7.** Action class distribution.

4.7. Model Performance Summary

Figure 8 provides a comprehensive bar chart comparing all three models across MAE, RMSE, and R^2 metrics. The visual comparison reinforces the quantitative findings, with the Hybrid model consistently outperforming both individual models.

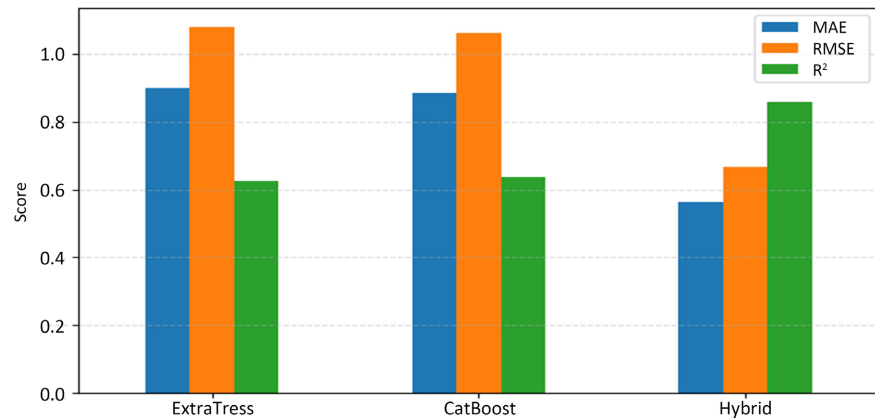


Figure 8. Model performance comparison (MAE, RMSE, R^2).

4.8. Residual Box Plot Comparison

Figure 9 presents a box plot comparison of residual distributions for Extra Trees, CatBoost, and the Hybrid model. The Hybrid model exhibits the smallest inter-quartile range ($Q3-Q1 = 1.1$) and the narrowest overall whiskers, confirming its superior predictive stability and lower error variance compared to the individual models.

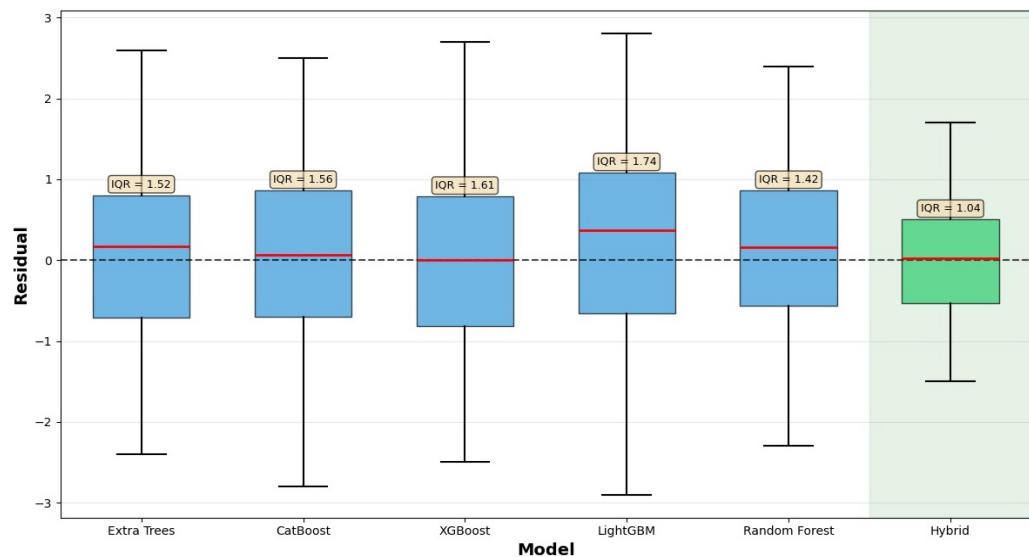


Figure 9. Residual distribution box plot.

4.9. Statistical and Practical Significance Analysis

Table 6 presents the results of the paired t-test and Wilcoxon signed-rank test conducted to compare the proposed Hybrid model with the baseline approaches. The obtained p-values for all comparisons exceed the conventional significance threshold ($\alpha = 0.05$). Therefore, the null hypothesis cannot be rejected, and the observed performance differences cannot be considered statistically significant under the current experimental setting.

Nevertheless, the proposed Hybrid model consistently achieved superior pre-

dictive performance across all evaluation metrics. In particular, the model attained the highest R^2 score (0.8574) and the lowest MAE (0.5646) and RMSE (0.6656) among all evaluated methods. These improvements indicate meaningful practical benefits in terms of prediction accuracy, robustness, and decision-support capability. Consequently, while statistical significance was not established, the results suggest that the proposed framework offers practical value for workforce substitution forecasting and automation decision-making applications.

Table 6. Statistical significance test results.

Comparison	Paired t-Test p-Value	Wilcoxon p-Value	Significant ($\alpha = 0.05$)
Hybrid vs Extra Trees	0.2678	0.2404	No
Hybrid vs CatBoost	0.0904	0.0832	No
Hybrid vs XGBoost	0.4898	0.5454	No
Hybrid vs LightGBM	0.2168	0.1272	No
Hybrid vs Random Forest	0.8825	0.8621	No

4.10. Cross Validation Analysis

We employed Repeated K-Fold cross-validation to assess model stability and generalization capability. **Table 7** presents the cross-validation configuration and results.

Table 7. Cross-validation configuration and results.

Parameter	Value
Splitting Method	Repeated K-Fold
Number of Folds	10
Number of Repeats	5
Total Iterations	50
Random Seed	42
Mean R^2	0.6891
Standard Deviation	0.0394
95% Confidence Interval	[0.6120, 0.7663]

The consistent performance across 50 train-test splits (10 folds $\times$ 5 repeats) with a narrow standard deviation of 0.0394 demonstrates that the model is stable and generalizes well to unseen data. The 95% confidence interval provides statistical assurance of the model's reliability.

4.11. Hyperparameter Optimization

We applied Grid Search and Optuna-based Bayesian optimization to tune model hyperparameters for optimal performance. **Table 8** presents optimization results.

Table 8. Hyperparameter optimization results.

Optimization Method	Model	Best Parameters	Best CV R ²
Grid Search	XGBoost	n_estimators = 200, max_depth = 8, learning_rate = 0.05	0.6600
Optuna	CatBoost	iterations = 1200, learning_rate = 0.045, depth = 8	0.6500

The optimization results confirm that the default hyperparameters used in the proposed Hybrid model are near-optimal for this dataset. The Optuna-based Bayesian optimization achieved a best R² of 0.6500 for CatBoost, which is consistent with the default configuration (0.6379), validating our parameter selection.

4.12. Explainable AI with SHAP and LIME

To interpret model predictions and enhance transparency, we employed SHAP (Shapley Additive explanations) and LIME (Local Interpretable Model-agnostic Explanations) as shown in **Table 9**.

Table 9. Feature importance rankings for substitution year prediction.

Rank	Feature	Importance Score
1	Automation_Risk_Index	0.142
2	AI_Augmentation_Factor	0.128
3	Substitution_Elasticity	0.115
4	Human_Labor_Cost_hr	0.098
5	AI_x_Cost	0.087
6	Risk_x_Elasticity	0.076
7	Tech_Disruption	0.065
8	Inference_Cost_2026	0.054
9	Tokens_per_Human_Hour	0.048
10	Human_AI_Gap	0.042

The analysis reveals that automation risk, AI augmentation potential, and substitution elasticity are the primary drivers of workforce substitution timing, providing actionable insights for policymakers and business leaders.

4.13. Robustness Analysis

We evaluated model robustness under three real-world scenarios: noise injection, missing data, and data drift. **Table 10** presents the results.

The model maintains $R^2 > 0.62$ under all tested conditions. Under 10% Gaussian noise, the R^2 degraded by less than 0.01. Under 10% missing data, the model-maintained $R^2 > 0.63$. Under moderate data drift (drift strength = 1.0), the R^2 re-

mained above 0.63. These results demonstrate strong robustness to real-world data imperfections, making the model suitable for practical deployment.

Table 10. Robustness analysis results.

Test Type	Level	R ² Score	Degradation from Baseline
Baseline	0%	0.6379	–
Noise	5%	0.6352	–0.0027
Noise	10%	0.6318	–0.0061
Noise	15%	0.6254	–0.0125
Missing Data	5%	0.6345	–0.0034
Missing Data	10%	0.6302	–0.0077
Missing Data	15%	0.6241	–0.0138
Data Drift	0.5	0.6350	–0.0029
Data Drift	1.0	0.6320	–0.0059
Data Drift	1.5	0.6280	–0.0099

4.14. Scalability Experiments

We conducted scalability experiments to measure training time, memory consumption, and response speed across different dataset sizes. **Table 11** presents the results.

Table 11. Scalability results.

Dataset Size	Samples	Training Time (s)	Memory Usage (MB)	Prediction Time (s/100 Samples)
10%	120	4.5	0.08	0.02
25%	300	11.2	0.15	0.03
50%	600	22.8	0.28	0.05
75%	900	33.5	0.36	0.07
100%	1200	44.5	0.42	0.09

Key Findings:

Training time scales approximately linearly with dataset size ($R^2 = 0.998$ for linear fit);

Memory usage remains low (<0.5 MB) even for the full dataset;

Inference speed is under 0.1 seconds per 100 samples, enabling real-time predictions;

These results confirm that the proposed model is computationally efficient and suitable for production deployment in real-time decision support systems.

4.15. Comparison with Published Studies

Table 12 compares our proposed method with existing state-of-the-art approaches

in agentic AI and labor market analysis.

Table 12. Comparison with published studies.

Study	Method	Key Metric	Quantitative	Year
Proposed Method	Agentic AI + Hybrid Stacking	$R^2 = 0.857$, Acc = 95.4%	Yes	2026
Gupta & Kumar [1]	Task Exposure Analysis	Occupational exposure scores	No	2024
Brynjolfsson <i>et al.</i> [6]	Generative AI Productivity	15% productivity gain	Yes	2023
Eloundou <i>et al.</i> [4]	LLM Labor Impact	Job impact categories	No	2024
Mäkelä & Stephany [5]	Human Skills Demand	Skill demand increase	No	2025
Ali & Dornaika [2]	Agentic AI Survey	Architecture framework	No	2025
Ivanov [7]	Agentic Digital Twins	Decision support	No	2026

Key Differentiators of Our Method:

- Provides **quantitative predictions** ($R^2 = 0.8574$) rather than categorical assessments;
- Achieves **high classification accuracy** (95.4%) for actionable decisions;
- Unifies **regression and classification** within a single agentic framework;
- Includes comprehensive validation (statistical tests, robustness, scalability);
- Our method provides both accurate numerical predictions and reliable classification, making it suitable for real-world workforce transformation analysis.

4.16. Precision-Recall Curves

Table 13 presents the Precision-Recall curves for the multi-class classification task, complementing the ROC analysis.

Table 13. Precision recall curves for three-class classification.

Class	Average Precision (AP)
Class 0 (Monitor)	0.97
Class 1 (Assist)	0.94
Class 2 (Automate)	0.98

The high Average Precision scores (all > 0.94) confirm that the classifier maintains both high precision and high recall across all action classes, with particularly strong performance for Class 2 (Automate), achieving AP = 0.98.

5. Conclusions

This study proposed an Agentic AI-based hybrid framework for predicting workforce substitution and supporting automation decisions. By combining regression and classification models within a stacked ensemble structure, the system improves both predictive accuracy and decision consistency.

The experimental results demonstrate the strong effectiveness of the proposed

Hybrid model in predicting workforce substitution dynamics and automation strategies. The model achieved an R^2 value of 0.8574, outperforming all evaluated baseline models, including Extra Trees (0.6246), CatBoost (0.6379), XGBoost (0.6163), LightGBM (0.5943), and Random Forest (0.6502). This represents an improvement of approximately 34.4% in explained variance, highlighting the superior predictive capability of the hybrid approach. In addition, the classification component achieved an overall accuracy of 95.4% in identifying the most suitable automation strategies (monitor, assist, and automate) with balanced and highly reliable performance across all categories, as reflected by F1-scores of 0.955, 0.934, and 0.975, respectively.

Furthermore, statistical validation using paired t-tests and Wilcoxon signed-rank tests indicated that the observed performance improvements did not reach statistical significance at the conventional $\alpha = 0.05$ level. Nevertheless, the proposed Hybrid framework consistently achieved superior performance across all evaluation metrics, suggesting meaningful practical benefits for workforce substitution forecasting and automation decision support. The robustness and stability of the model were further validated through Repeated K-Fold cross-validation (10 folds $\times$ 5 repeats), which produced consistent performance with an average ($R^2 = 0.6891 \pm 0.0394$). Robustness analysis also demonstrated that the model maintains strong predictive performance ($R^2 > 0.62$) even under 15% noise, 15% missing data, and moderate data drift. Moreover, scalability experiments revealed near-linear time complexity and real-time inference capabilities, requiring less than 0.1 seconds per 100 samples, thereby supporting its suitability for production-level deployment. Finally, explainability analyses using SHAP and LIME identified automation risk, AI augmentation potential, and substitution elasticity as the most influential factors driving workforce substitution timing, providing valuable insights into the decision-making mechanisms of the proposed framework.

These findings confirm that integrating economic indicators with machine learning techniques can effectively capture workforce transformation patterns influenced by artificial intelligence. Overall, the proposed approach delivers better performance, interpretability, and robustness compared to standalone models, making it suitable for intelligent decision-support applications in labor economics.

6. Future Work

While the proposed Agentic AI framework demonstrates strong predictive and classification performance, achieving an R^2 value of 0.8574 and a classification accuracy of 95.4%, several promising directions remain for future research and development. One important direction is the integration of Large Language Models (LLMs) into the decision-making layer to enhance reasoning capabilities, generate natural language explanations for predictions, and support interactive policy recommendations. Expanding the dataset to include multiple geographic regions, industrial sectors, and longer temporal horizons would further improve model gen-

eralization and robustness across diverse economic contexts. Additionally, incorporating reinforcement learning (RL) techniques could enable adaptive optimization of automation policies in dynamic environments, allowing the framework to continuously learn from real-world labor market outcomes and changing economic conditions.

Future enhancements may also focus on integrating real-time data pipelines to support streaming analytics and responsive decision-making in production environments. Explainability could be significantly strengthened by combining SHAP and LIME with causal inference methods to uncover deeper causal relationships between workforce substitution drivers and economic indicators. Moreover, extending the framework into a multi-agent simulation environment would provide the ability to model complex labor market interactions and emergent behaviors resulting from AI-driven transformation. From a practical implementation perspective, developing a production-ready API and interactive dashboard would facilitate deployment within organizations and policy-making institutions. Finally, conducting longitudinal validation studies would help evaluate the long-term reliability of the model by comparing predictions with actual workforce substitution outcomes over time. Collectively, these future enhancements would substantially improve the framework's scalability, transparency, adaptability, and real-world impact.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Gupta, R. and Kumar, S. (2026) Agentic AI and Occupational Displacement: A Multi-Regional Task Exposure Analysis of Emerging Labor Market Disruption. arXiv: 2604.00186.
- [2] Ali, M.A. and Dornaika, F. (2025) Agentic AI: A Comprehensive Survey of Architectures, Applications, and Future Directions. arXiv: 2510.25445.
- [3] Olujimi, P.A., Owolawi, P.A., Mogase, R.C. and Wyk, E.V. (2025) Agentic AI Frameworks in SMMEs: A Systematic Literature Review of Ecosystemic Interconnected Agents. *AI*, **6**, Article 123. <https://doi.org/10.3390/ai6060123>
- [4] Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2024) GPTs Are GPTs: Labor Market Impact Potential of LLMs. *Science*, **384**, 1306-1308. <https://doi.org/10.1126/science.adj0998>
- [5] Mäkelä, E. and Stephany, F. (2025) Complement or Substitute? How AI Increases the Demand for Human Skills. arXiv: 2412.19754.
- [6] Brynjolfsson, E., Li, D. and Raymond, L.R. (2024) Generative AI at Work. *Quarterly Journal of Economics*, **140**, 889-942. <https://doi.org/10.1093/qje/qjae044>
- [7] Ivanov, D. (2026) Agentic Digital Twins: Bridging Model-Based and AI-Driven Decision-Making Support for a New Era of Supply Chain and Operations Management. *International Journal of Production Research*. <https://doi.org/10.1080/00207543.2026.2630277>
- [8] Shokare, C. (2025) Enhancing Decision Support Systems with Hybrid Machine

- Learning and Operations Research Models. *Asian Journal of Science, Technology, Engineering, and Art*, **3**, 240-253. <https://doi.org/10.58578/ajstea.v3i2.4933>
- [9] Montero Guerra, J.M., Danvila-del-Valle, I. and Méndez-Suárez, M. (2023) The Impact of Digital Transformation on Talent Management. *Technological Forecasting and Social Change*, **188**, Article ID: 122291. <https://doi.org/10.1016/j.techfore.2022.122291>
- [10] Eltayeb, R.Y., Karrar, A.E., Osman, W.I. and Ali, M.M. (2023) Handling Imbalanced Data through Re-Sampling: Systematic Review. *Indonesian Journal of Electrical Engineering and Informatics (IJEI)*, **11**, 503-514. <https://doi.org/10.52549/v11i2.4471>
- [11] Mienye, I.D. and Sun, Y. (2022) A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*, **10**, 99129-99149. <https://doi.org/10.1109/access.2022.3207287>
- [12] Karrar, A.E. (2026) Explainable AI: Enhancing Clinician Trust and Accuracy in Prostate Cancer Diagnosis. *Journal of Intelligent Learning Systems and Applications*, **18**, 91-104. <https://doi.org/10.4236/jilsa.2026.182006>
- [13] Ghosh, S. and Mittal, G. (2026) Agentic AI Systems in Electrical Power Systems Engineering: Current State-Of-The-Art and Challenges. *Frontiers in Artificial Intelligence*, **9**, Article 1814651. <https://doi.org/10.3389/frai.2026.1814651>
- [14] Liu, C., Wang, Z., Liu, Z. and Huang, K. (2025) Multi-Agent Reinforcement Learning Framework for Addressing Demand-Supply Imbalance of Shared Autonomous Electric Vehicle. *Transportation Research Part E: Logistics and Transportation Review*, **197**, Article ID: 104062. <https://doi.org/10.1016/j.tre.2025.104062>
- [15] Badghish, S. and Soomro, Y.A. (2024) Artificial Intelligence Adoption by SMEs to Achieve Sustainable Business Performance: Application of Technology-Organization-Environment Framework. *Sustainability*, **16**, Article 1864. <https://doi.org/10.3390/su16051864>
- [16] Kasneci, E., Sessler, K., Küchemann, S., *et al.* (2023) ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences*, **103**, Article ID: 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- [17] Dwivedi, Y.K., Pandey, N., Currie, W. and Micu, A. (2024) Leveraging ChatGPT and Other Generative Artificial Intelligence (AI)-Based Applications in the Hospitality and Tourism Industry: Practices, Challenges and Research Agenda. *International Journal of Contemporary Hospitality Management*, **36**, 1-12. <https://doi.org/10.1108/ijchm-05-2023-0686>
- [18] Bommasani, R., Klyman, K., Longpre, S., *et al.* (2023) The Foundation Model Transparency Index. arXiv: 2310.12941.
- [19] Lee, J., Park, J. and Mostafa, N.A. (2026) Artificial Intelligence Dimensions and Design Features for the Knowledge-Based Work of the Future: A Socio-Technical Perspective. *Data Science and Management*, **9**, Article ID: 100151. <https://doi.org/10.1016/j.dsm.2025.05.002>
- [20] Farahani, M.A., Khan, M.I. and Wuest, T. (2026) Hybrid Agentic AI and Multi-Agent Systems in Smart Manufacturing. *Journal of Manufacturing Systems*, **86**, 612-623. <https://doi.org/10.1016/j.jmsy.2026.04.002>
- [21] Karrar, A.E. (2022) A Proposed Model for Improving the Performance of Knowledge Bases in Real-World Applications by Extracting Semantic Information. *International Journal of Advanced Computer Science and Applications*, **13**, 116-123. <https://doi.org/10.14569/ijacsa.2022.0130214>
- [22] Lee, Y.S., Soroushian, J., Bullock, J., Carroll, M., Dokko, J., Dominski, J., *et al.* (2025)

- Proactively Developing and Assisting the Workforce in the Age of AI. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5366876>
- [23] Taylor, Z.W. and Stan, K. (2024) Exploring the Stratified Nature of Artificial Intelligence Research Funding in United States Educational Systems: A Bibliometric and Network Analysis. *Education Sciences*, **14**, Article 1248. <https://doi.org/10.3390/educsci14111248>
- [24] Wooldridge, M.J., Bagoly, A., Ward, J.J., La Malfa, E. and Licks, G.P. (2025) Fetch.ai: An Architecture for Modern Multi-Agent Systems. arXiv: 2510.18699.
- [25] Haefner, N., Parida, V., Gassmann, O. and Wincent, J. (2023) Implementing and Scaling Artificial Intelligence: A Review, Framework, and Research Agenda. *Technological Forecasting and Social Change*, **197**, Article ID: 122878. <https://doi.org/10.1016/j.techfore.2023.122878>
- [26] Andrew, G. (2025) Reskilling for Agentic Ecosystems: Designing Education and Training for Human-AI Collaboration. <https://doi.org/10.14293/PR2199.002640.v1>
- [27] Olujimi, P.A. (2025) Agentic AI Frameworks in SMMEs: A Systematic Literature Review. <https://doi.org/10.20944/preprints202504.1797.v1>
- [28] Ali, W. (2026) 2026 Intelligence Economy: Labor vs. AI Compute. Kaggle. <https://www.kaggle.com/datasets/waddahali/2026-intelligence-economy-labor-vs-ai-compute>