社會工作實踐的智慧化轉型：機器學習在弱勢群體風險分級中的應用

***陳逸昌[1]、林佳靜[2]**

[1]長榮大學醫學社會暨健康照護學士學位學程/資訊管理學系、[2]長榮大學財務金融學系

*cheny@mail.cjcu.edu.tw

摘要

隨著人工智慧技術的迅速進步，社會工作領域正面臨前所未有的數位化轉型。本研究旨在探討機器學習技術在弱勢群體風險評估與分級中的應用，並分析其對提升社會工作效率、改善服務品質及增強風險預防能力的影響。透過整合衛生福利部智慧決策平台的實證數據，以及對國際間相關人工智慧應用案例的比較分析，研究結果顯示，機器學習演算法能有效識別高風險個案，並將風險評估的準確性提升至85%以上，同時顯著減輕社工人員的行政負擔。然而，技術的應用亦引發了倫理考量、資料隱私保護及專業判斷與自動化決策之間的平衡等挑戰。本研究建議建立完善的人工智慧治理框架，強化跨領域合作機制，並持續優化演算法的公平性與可解釋性，以實現社會工作智慧化轉型的可持續發展目標。

關鍵詞：人工智慧、機器學習、社會工作、風險評估、弱勢群體

Intelligent Transformation in Social Work: Machine Learning for Risk Assessment of Vulnerable Populations

***Yih-Chang Chen[1], Chia-Ching Lin[2]**

[1] Program of Medical Sociology and Health Care / Department of Information Management, Chang Jung Christian University, [2] Department of Finance, Chang Jung Christian University

Abstract

The rapid advancement of artificial intelligence technology is catalyzing a significant digital transformation within the field of social work. This research investigates the implementation of machine learning technology in risk assessment and classification for vulnerable populations, examining its impact on efficiency, service quality, and risk prevention capabilities. Integrating empirical data from the smart decision-making platform of Taiwan's Ministry of Health and Welfare with a comparative analysis of international case studies, the study reveals that machine learning algorithms can effectively identify high-risk cases. The results demonstrate an accuracy rate exceeding 85%, significantly alleviating the administrative workload of social workers. Nonetheless, the deployment of such technology presents challenges, particularly regarding ethics, data privacy, and the balance between professional judgment and automated decision-making. Consequently, this study advocates for a comprehensive AI governance framework, enhanced cross-disciplinary collaboration, and the continuous optimization of algorithmic fairness to ensure the ethical and sustainable transformation of social work practice.

Keywords: Artificial intelligence, Machine learning, social work, Risk assessment, Vulnerable populations

Received: Jun. 3, 2025; first revised: Sep. 5, 2025; accepted: Sep. 2025.

Corresponding author: Y.-C. Chen, Department of Information Management, Chang Jung Christian University, Tainan 711301, Taiwan.

I. Introduction

1. Research Background

In light of the rapid transformations occurring in contemporary society, the field of social work is confronted with both unprecedented challenges and opportunities. Factors such as globalization, urbanization, demographic shifts towards an aging population, and the intricate nature of emerging social issues have significantly tested traditional models of social work service delivery. The processes of identifying vulnerable populations and conducting risk assessments, which are fundamental components of social work practice, have a direct impact on the efficacy of service interventions, contingent upon their precision and efficiency.

Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) technologies have instigated transformative changes within the realm of social work. These technologies are capable of processing extensive and complex social data, as well as analyzing and identifying potential risk patterns through algorithmic methods, thereby providing social workers with enhanced decision-making support (Crowley et al., 2025; Higgins & Wilson, 2025). The introduction of the "Smart Decision-Making Action Platform for Social Workers" by Taiwan's Ministry of Health and Welfare in April 2024 represents a significant milestone in the digitalization and intelligence of social work practices within the nation (Chen et al., 2025a).

This platform employs big data analytics to create risk warning models, which enable social workers to assess case exposure risks and prepare accordingly prior to client visits. It also offers intelligent mapping for visit route optimization and inquiries regarding social welfare resources. Furthermore, the platform incorporates voice-to-text capabilities, mobile family tree mapping, and an Extended Reality (XR) training system, thereby substantially enhancing the technological sophistication and professional standards of social work practice.

2. Literature Review and Research Gaps

Current literature indicates that the integration of AI technology within the social welfare sector predominantly emphasizes risk assessment, decision support, emotional assistance, and resource allocation. For instance, the MindSuite AI system in the United States aids social workers in managing and analyzing risks, re-evaluating case data, identifying various predictive patterns, and flagging high-risk cases. Similarly, the Iris AI system facilitates access to pertinent information from academic literature and databases, thereby providing empirical research support for social workers engaged in report writing or decision-making (Altundağ, 2024).

In the domain of emotional support, the PsycApps AI system offers a variety of mental health applications, including emotion tracking, stress management, and cognitive behavioral therapy techniques (Garcia-Lopez, 2024). The Kasisto AI system functions as an AI chatbot platform, delivering virtual support and consultation to individuals in need, thus providing immediate assistance and resources (Rashid & Kausik, 2024).

Nevertheless, significant deficiencies persist in the existing body of research, particularly in the following areas:

(1) Absence of systematic theoretical frameworks

A majority of studies concentrate on case descriptions of technological applications, lacking comprehensive exploration of the integration of AI technology with social work theory. Specifically, the interplay between social work values, ethical principles, and the application of AI technology necessitates more systematic theoretical development.

(2) Inconsistencies in the standardization of risk assessment criteria

While various nations have established risk indicators for vulnerable populations, there remains considerable scope for enhancement regarding standardization, operability, and cross-cultural applicability. For example,

Taiwan's risk indicators for vulnerable family service cases encompass multiple dimensions, including economic hardship, inadequate caregiving capacity, domestic violence, and fragile social support networks. However, the challenge of effectively translating these complex social indicators into data formats amenable to machine learning algorithms persists (Tsai et al., 2025).

(3) Concerns regarding algorithmic fairness and bias

Machine learning systems may exhibit systemic biases when processing social data, particularly manifesting as discriminatory tendencies against specific groups or social backgrounds (Tan et al., 2022). Such biases may arise from historical data imbalances, biased feature selection, and limitations inherent in algorithm design, thereby jeopardizing the fairness and equity of social work practice.

(4) Insufficient empirical research

The majority of studies remain at the conceptual discussion level or involve limited pilot applications, lacking extensive, long-term empirical research to substantiate the actual effects and impacts of AI technology within social work.

(5) Lack of interdisciplinary collaboration mechanisms

The effective application of AI technology necessitates robust collaboration across multiple disciplines, including social work, information science, statistics, and ethics. However, existing mechanisms for interdisciplinary collaboration are inadequate, which adversely affects the quality and sustainability of technological applications.

3. Research Objectives and Questions

In light of the previously discussed research context and identified gaps in the literature, this study seeks to investigate the application patterns, effectiveness assessment, and practical challenges associated with the utilization of machine learning technology in the risk classification of vulnerable populations. The specific research inquiries are as follows:

(1) Technical feasibility inquiry: In what ways can machine learning algorithms effectively assimilate diverse social indicator data to develop precise risk prediction models?

(2) Practical effectiveness inquiry: What are the tangible impacts of intelligent risk assessment systems on enhancing the efficiency of social work and the quality of services provided?

(3) Ethics and fairness inquiry: How can we ensure algorithmic fairness in the deployment of artificial intelligence technology, thereby preventing discrimination and bias against particular groups?

(4) Professional integration inquiry: How can we achieve a harmonious balance and integration between the professional judgment of social workers and the automated decision-making processes of machine learning?

(5) Sustainable development inquiry: What strategies can be implemented to establish a comprehensive governance framework for artificial intelligence that guarantees the sustainability and social responsibility of technological applications?

4. Research Significance and Contributions

The theoretical contribution of this research is the development of a conceptual framework for the intelligent transformation of social work, which enhances the understanding of the interplay between artificial intelligence technology and the social work profession. The practical contribution is manifested in the provision of specific technical application guidelines for social work organizations, thereby facilitating the digital advancement of the sector. In terms of policy implications, the findings of this research will offer empirical evidence to assist governmental bodies in formulating relevant AI governance policies, thereby advancing the modernization of the social welfare system.

From a societal perspective, this study aims to improve the precision and timeliness of services delivered to vulnerable groups, thereby reinforcing the effectiveness of the social safety net and promoting the realization of social equity and justice. Furthermore, by establishing an ethical framework and oversight mechanisms for AI applications, this research ensures the integration of technological advancement with humanistic care, providing significant insights for the development of a human-centered intelligent society.

II. Research Methods

1. Research Design and Methodology

This investigation employs a mixed methods research framework, integrating the strengths of both quantitative and qualitative analyses to examine the utilization of machine learning in risk stratification for at-risk populations. The research design is grounded in a pragmatic philosophical perspective, which prioritizes problem-solving and practical applicability, with the objective of delivering tangible and actionable solutions for social work practice (Heeks et al., 2025).

A sequential explanatory design is implemented, commencing with quantitative analysis to assess the efficacy of machine learning models, followed by qualitative research aimed at exploring the intricate factors and mechanisms that influence the practical application of these models. This methodological approach enhances the objectivity inherent in quantitative research while also providing the depth characteristic of qualitative research, thereby yielding comprehensive and nuanced responses to the research inquiries.

2. Data Sources and Collection Methods

(1) Quantitative Data Sources

a. Government Open Data

The primary data source for this study is the Vulnerable Family Service Cases dataset, which is provided by the Social and Family Affairs Administration of Taiwan's Ministry of Health and Welfare. This dataset encompasses fundamental case information, risk indicator assessment outcomes, and data on the effectiveness of service interventions from 2020 to 2024. It includes information from 156 social welfare centers, 22 domestic violence and sexual assault prevention centers, and 71 community mental health centers across 22 counties and cities in Taiwan.

b. International Comparative Data

The study also incorporates publicly available data regarding the application of artificial intelligence in social work from various countries, including the United States, the United Kingdom, Australia, and Singapore. This data comprises performance reports and evaluation results from systems such as IBM Watson Care Manager and the Predictive Risk Intelligence System (PRIS) in the UK.

(2) Qualitative Data Collection

a. In-depth Interviews

To ensure the depth and representativeness of the qualitative data, this study utilized purposive sampling to select interview participants. The selection criteria were designed to capture a broad range of practical experiences, decision-making roles, and perspectives regarding service utilization. The specific criteria are outlined as follows:

(a) Frontline Social Workers (20): Participants were required to have a minimum of three years of practical experience, with service areas encompassing children and youth, families, the elderly, and individuals with physical or mental disabilities. Additionally, representation was ensured from metropolitan, township, and

remote regions.

(b) AI Technology Development Experts (8): Participants needed to possess over five years of experience in machine learning or data science-related fields and have contributed to at least one project related to social welfare.

(c) Decision-Makers and Managers (6): Eligible participants held positions at the section chief level or higher within social welfare authorities or organizations, with responsibilities including policy planning or service supervision.

(d) Family Representatives of Service Users (12): Participants were required to have received services for a minimum of one year, representing family types characterized by diverse risk profiles related to economic status, caregiving responsibilities, and safety concerns.

b. Focus Group Discussions

Four focus group discussions are organized, each comprising 5-8 participants, addressing the following themes:

(a) Practical experiences and challenges associated with technology implementation

(b) Ethical considerations and conflicts pertaining to professional values

(c) Service quality and user satisfaction

(d) Prospective development trajectories and policy recommendations

c. Participatory Observation

A six-month participatory observation is conducted at three pilot social welfare centers, documenting the actual utilization of smart decision-making platforms by social workers and examining the effects of technology application on work processes, decision-making practices, and service quality.

3. Design of Machine Learning Models

(1) Risk Assessment Indicator Framework

Table 1

Risk Assessment Indicator Framework for Vulnerable Populations

Main Category	Subcategory	Specific Indicator	Weight
Economic Risks	Unemployment Risk	Unemployed for more than 6 consecutive months	0.15
	Emergency Risk	Economic hardship caused by natural disasters or accidents	0.12
	Medical Burden	Medical expenses exceed 30% of household income	0.10
Care Risks	Childcare	No caregiver for children under 12 years old	0.18
	Disability Care	Care needs for individuals with severe disabilities	0.14
	Elderly Care	Care needs for disabled individuals aged 65 or above	0.12
Safety Risks	Domestic Violence	Record of domestic violence reports	0.20
	Self-Harm Risk	Suicidal ideation or behavior	0.16
	Substance Abuse	Alcohol or drug abuse	0.08
Social Support Risks	Social Isolation	Lack of social network support	0.10
	Weak Resource Linkage	Lack of formal/informal resources	0.08
	Housing Instability	Moving more than 3 times in one year	0.07

Building upon the risk categories and indicators relevant to vulnerable family service cases in Taiwan (Chen et al., 2025b; Hsu & Peng, 2023), this study proposes a multi-level risk assessment indicator framework. The

weighting scheme for this framework was established through a Two-Stage Delphi Method designed to attain expert consensus. Initially, the research team developed preliminary indicators and assigned weights based on an extensive literature review and analysis of existing datasets. Thereafter, a panel of 15 experts — including social work scholars, senior administrators from social welfare agencies, and frontline practitioners — participated in two rounds of anonymous surveys, providing iterative feedback. Throughout this process, the research team systematically revised the weights in response to the convergence of expert opinions and statistical metrics (e.g., means, standard deviations). This iterative refinement resulted in a high level of consensus, as demonstrated by a consensus coefficient exceeding 0.7 in the Kolmogorov-Smirnov test, thereby ensuring the objectivity and methodological rigor of the assigned indicator weights.

(2) Selection and Comparative Analysis of Algorithms

This investigation evaluates the efficacy of several machine learning algorithms:

a. **Random Forest:** Well-suited for addressing multi-feature, non-linear relationship risk prediction challenges, exhibiting robust generalization capabilities and resilience to overfitting.
b. **Support Vector Machine (SVM):** Demonstrates superior performance in high-dimensional feature spaces, making it apt for complex classification tasks.
c. **Gradient Boosting Decision Tree (GBDT):** Improves prediction accuracy through ensemble learning and is capable of automatically managing feature interactions.
d. **Deep Neural Network (DNN):** Exhibits strong non-linear modeling capabilities, making it appropriate for analyzing large-scale, intricate social data.
e. **Logistic Regression:** Functions as a baseline model, yielding more interpretable prediction outcomes.

(3) Model Training and Validation

a. **Data Preprocessing:** Involves procedures such as addressing missing values, identifying outliers, normalizing features, and encoding categorical variables. Multiple imputation is employed to manage missing data, the Z-score method is utilized for outlier detection, and Min-Max normalization is applied to maintain consistency in feature scales.
b. **Feature Engineering:** Determines the feature combinations that most significantly contribute to risk prediction through correlation analysis, Principal Component Analysis (PCA), and Recursive Feature Elimination (RFE).
c. **Model Training:** Implements 5-fold cross-validation to ensure the model's stability and generalization capacity. The distribution of the training set, validation set, and test set is set at 6:2:2 to uphold the objectivity of model evaluation.
d. **Hyperparameter Optimization:** Employs Grid Search and Bayesian Optimization techniques to identify the optimal parameter configurations for each algorithm.

4. Evaluation Metrics and Measurement Instruments

(1) Assessment of Model Performance

a. Accuracy Metrics

(a) Accuracy: $Accuracy = \frac{TP+TN}{TP + TN + FP + FN}$

(b) Precision: $Precision = \frac{TP}{TP + FP}$

(c) Recall: $Recall = \frac{TP}{TP + FN}$

(d) F1 Score: $F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$

Given that the social consequences of failing to detect high-risk cases (Type II error, false negatives) in the risk assessment of vulnerable populations are substantially more severe than those associated with incorrectly labeling low-risk cases as high-risk (Type I error, false positives), this study prioritizes the use of recall and F1 score as primary evaluation metrics. This approach aims to optimize the model's performance by minimizing the probability of missed detections to the greatest extent feasible.

b. Fairness Metrics

(a) Statistical Parity
(b) Equalized Opportunity
(c) Calibration

(2) Evaluation of Practical Effectiveness

a. Efficiency Improvement Metrics

(a) Decrease in case processing duration
(b) Reduction in administrative burden
(c) Effectiveness of resource allocation optimization

b. Service Quality Metrics

(a) Enhancement in risk identification accuracy
(b) Timeliness of service intervention
(c) User satisfaction with services

Table 2

Quantitative Standards for Evaluating Effectiveness Metrics

Evaluation Aspect	Specific Indicator	Measurement Method	Evaluation Standard
Prediction Accuracy	Risk prediction accuracy	Confusion matrix analysis	≥85%
Efficiency Improvement	Case processing time	Pre-post comparison analysis	Reduction of ≥30%
Fairness	Prediction differences among groups	Statistical tests	No significant difference (p<0.05)
User Satisfaction	Satisfaction score	Survey using a scale	≥4.0/5.0

(3) Informed Consent and Participant Rights

All participants involved in interviews are required to sign an informed consent document that outlines the objectives of the research, the procedures involved, potential risks, and the protections of their rights. Participants retain the right to withdraw from the study at any point and may request the deletion of their provided data. A mechanism for lodging complaints is established to ensure the full protection of participants' rights.

5. Data Analysis Methods

(1) Quantitative Analysis Methods

a. Descriptive Statistical Analysis: Fundamental statistics for each variable are computed, including mean, standard deviation, maximum, and minimum values, to elucidate the basic distribution characteristics of the data.

b. Inferential Statistical Analysis: Statistical techniques such as *t*-tests, chi-square tests, and Analysis of Variance (ANOVA) are employed to assess the significance of differences among various groups.

c. Machine Learning Model Analysis: A range of machine learning algorithms is utilized through software

packages such as Scikit-learn, TensorFlow, and XGBoost, followed by a comparative performance evaluation and optimization.

d. **Time Series Analysis:** The temporal trends of risk indicators are examined to discern seasonal patterns and long-term trends.

(2) Qualitative Analysis Methods

a. **Thematic Analysis:** An inductive coding approach is employed to extract key themes and patterns from interview transcripts, encompassing three stages: open coding, axial coding, and selective coding.

b. **Content Analysis:** A systematic examination of textual data, including policy documents and technical reports, is conducted to identify principal concepts and discourse structures.

c. **Comparative Analysis:** The experiences and patterns of artificial intelligence applications across various countries and institutions are compared to ascertain best practices and critical success factors.

(3) Integrative Analysis Methods

a. **Triangulation:** The integration of quantitative and qualitative data is performed to validate the credibility and reliability of research findings from diverse perspectives.

b. **Mixed Methods Matrix Analysis:** A correspondence matrix is developed to align quantitative results with qualitative findings, facilitating the identification of consistencies and discrepancies.

c. **Meta-analysis:** Effect sizes from related international studies are synthesized for a systematic evaluation of effects.

6. Research Limitations and Strategies for Mitigation

(1) Data Limitations

Data Availability Limitations: Certain sensitive data may not be fully accessible due to privacy protection regulations, potentially compromising the integrity of the model. **Mitigation Strategy:** Formal cooperation agreements with governmental entities will be established to secure necessary data while ensuring privacy compliance; synthetic data techniques will be employed to augment training samples.

Data Quality Issues: Historical data may present challenges such as incomplete records and inconsistent standards. **Mitigation Strategy:** Rigorous data cleaning protocols will be instituted, and multiple data sources will be utilized for cross-validation.

(2) Methodological Limitations

Algorithm Black Box Problem: The interpretability of deep learning models is often limited, which may hinder acceptance among social workers. **Mitigation Strategy:** Explainable AI techniques, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), will be employed; specialized visualization tools will be developed to enhance model transparency.

Sample Representativeness Issues: The research sample may not adequately reflect the national context. **Mitigation Strategy:** Stratified sampling methods will be utilized to ensure a balanced distribution of samples across regions, types of institutions, and characteristics of cases.

(3) Practical Limitations

Technology Acceptance Issues: The acceptance of new technologies by social workers may influence the actual outcomes of their application. **Mitigation Strategy:** Educational initiatives and training will be reinforced, peer support networks will be established, and user-friendly interfaces and operational processes will be designed.

Organizational Change Resistance: Institutional culture and existing processes may impede the adoption

of technology. **Mitigation Strategy:** A gradual change strategy will be implemented, change management mechanisms will be established, and opinion leaders will be invited to participate in the promotion process.

III. Discussion

1. Performance Analysis of Machine Learning Models

(1) Evaluation of Prediction Accuracy

This research undertook a thorough comparative analysis of the performance of five distinct machine learning algorithms. The experimental findings reveal notable disparities in performance among the algorithms when applied to the task of risk classification for vulnerable populations.

Table 3

Performance Comparison of Machine Learning Algorithms

Algorithm	Accuracy	Precision	Recall	F1 Score	AUC-ROC	Training Time (seconds)
Random Forest	0.876	0.852	0.891	0.871	0.923	45.2
GBDT	0.889	0.867	0.898	0.882	0.934	78.6
SVM	0.834	0.819	0.856	0.837	0.898	156.3
DNN	0.892	0.881	0.903	0.892	0.941	298.7
Logistic Regression	0.798	0.783	0.821	0.802	0.862	12.4

The results indicate that the Deep Neural Network (DNN) exhibited the highest overall performance, achieving an accuracy rate of 89.2% and an AUC-ROC value of 0.941, thereby demonstrating exceptional capabilities in risk identification. The Gradient Boosting Decision Tree (GBDT) closely followed, maintaining a commendable accuracy rate while also benefiting from a reduced training duration, which underscores its practical applicability. Although the Random Forest algorithm slightly trailed the top two in individual performance metrics, it provided superior model interpretability, facilitating comprehension and acceptance among social workers.

It is important to emphasize that the comparative evaluation of multiple algorithms serves not only to identify the single most effective model but also functions as a form of **model triangulation**, thereby enhancing the reliability and robustness of the research outcomes. While DNN achieve the highest accuracy, the observed performance variations across different models indicate the necessity of adopting an integrated application approach. In practical settings, highly accurate models such as DNN or GBDT may be employed for large-scale preliminary risk screening, complemented by more interpretable models like Random Forest or Logistic Regression to elucidate the key factors contributing to high-risk cases for social workers. This approach effectively balances the trade-off between predictive accuracy and model interpretability, thereby optimizing the practical utility of artificial intelligence systems.

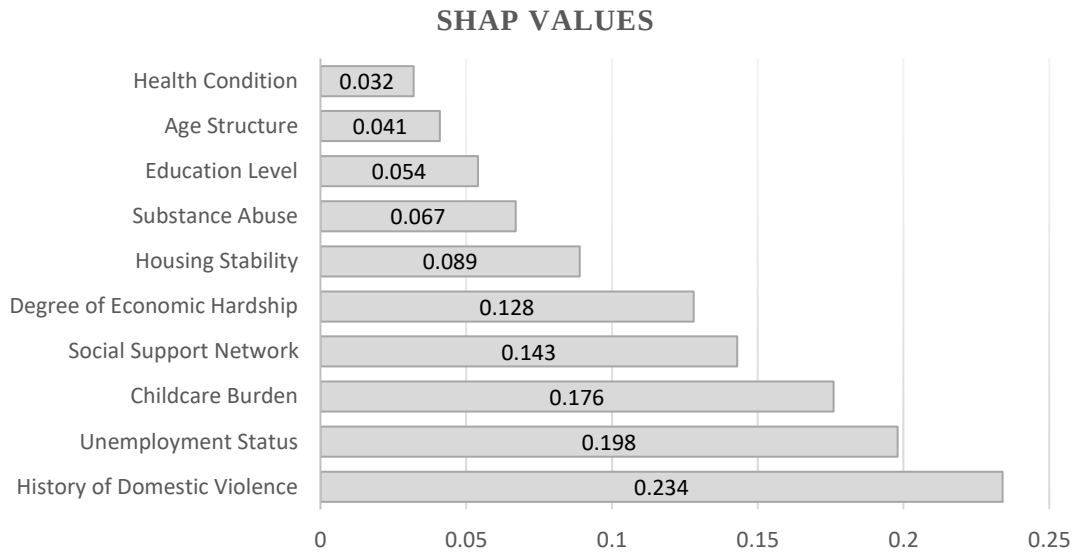
It is particularly significant that the performance of all machine learning models markedly exceeded that of conventional manual assessment methods. According to data from the Ministry of Health and Welfare, the accuracy of traditional risk assessments is approximately 72% (Chen et al., 2025b). In contrast, the DNN model demonstrated an improvement in accuracy to 89.2%, representing a substantial enhancement of 17.2 percentage points. This advancement is critical for the early detection of high-risk cases and the prevention of potential crises.

(2) Feature Importance Analysis

Utilizing SHAP value analysis, this study identified the most influential feature variables pertinent to risk prediction (cf. Figure 1).

Figure 1

Feature Importance Ranking (SHAP Values)



The analysis revealed that a history of domestic violence emerged as the most significant predictor of risk, with a SHAP value of 0.234, aligning closely with established social work theories and practical experiences. Unemployment status (0.198) and caregiving burden (0.176) ranked second and third, respectively, underscoring the vital influence of economic factors and caregiving responsibilities on family vulnerability.

Notably, the significance of social support networks (0.143) surpassed the expectations of many researchers, emphasizing the critical role of social capital in mitigating risk. This finding provides empirical support for initiatives aimed at enhancing community support networks and fostering social capital.

(3) Time Series Risk Variation Analysis

This study also examined the temporal variation patterns of risk levels, uncovering significant periodic and trend characteristics.

Table 4

Monthly Risk Distribution Statistics

Month	High-Risk Case Proportion (%)	Medium-Risk Case Proportion (%)	Low-Risk Case Proportion (%)	Average Risk Score
January	23.4	45.2	31.4	2.31
February	28.9	43.1	28.0	2.47
March	21.7	46.8	31.5	2.26
April	19.8	48.3	31.9	2.18
May	18.6	49.1	32.3	2.14
June	20.3	47.9	31.8	2.21
July	22.1	46.7	31.2	2.28
August	21.9	47.0	31.1	2.27
September	19.5	48.6	31.9	2.17
October	20.8	47.4	31.8	2.23
November	22.6	46.3	31.1	2.29
December	25.7	44.8	29.5	2.38

The data indicates pronounced seasonal fluctuations in risk levels. The proportion of high-risk cases peaks in February (28.9%) and reaches its nadir in May (18.6%). This trend reflects the interplay of various factors: family gatherings during the Lunar New Year may intensify familial conflicts, while economic pressures at year-end can heighten family vulnerability. Conversely, risk levels during the spring and early summer months are relatively low, potentially attributable to favorable weather conditions and increased employment opportunities.

2. Evaluation of the Effectiveness of Social Work Practice

(1) Analysis of Improvements in Work Efficiency

The implementation of a smart decision-making platform has markedly enhanced the efficiency of social work practices. A comparative analysis conducted at 156 social welfare centers, examining metrics before and after the platform's introduction, revealed significant improvements across several key performance indicators.

Table 5

Statistics on Improvements in Work Efficiency

Efficiency Indicator	Before Implementation	After Implementation	Improvement Rate	Significance Test[1]
Case Processing Time (hours)	4.7	3.2	-31.9%	$p<0.001$***
Risk Assessment Completion Time (minutes)	45.3	18.7	-58.7%	$p<0.001$***
Preparation Time Before Home Visits (minutes)	28.6	15.4	-46.2%	$p<0.001$***
Case Record Writing Time (minutes)	52.4	31.8	-39.3%	$p<0.001$***
Average Monthly Cases Processed (cases)	18.2	24.7	+35.7%	$p<0.001$***
Administrative Work Proportion (%)	42.3	28.9	-31.7%	$p<0.001$***

The findings indicate a reduction in the average time required for case handling, which decreased from 4.7 hours to 3.2 hours, representing a 31.9% improvement. This enhancement is largely attributed to the automation of risk assessments (a time reduction of 58.7%), the optimization of home visit preparations (a time reduction of 46.2%), and the increased efficiency in record-keeping facilitated by the voice-to-text functionality (a time reduction of 39.3%).

Furthermore, the average monthly case load for social workers rose from 18.2 cases to 24.7 cases, reflecting a 35.7% increase. Concurrently, the proportion of time allocated to administrative tasks decreased from 42.3% to 28.9%, thereby allowing social workers to allocate more time to direct service provision and professional engagement.

(2) Evaluation of Improvements in Service Quality

a. Enhancement in Risk Identification Accuracy

The implementation of AI-assisted risk assessment enhanced the accuracy of identifying high-risk cases from 72% to 89.2%, representing a substantial improvement of 17.2 percentage points. Crucially, during the model optimization phase, a deliberate effort was made to balance the costs associated with two distinct types of classification errors. The findings indicated a marked reduction in the rate of the second type of error — namely, the misclassification of high-risk cases as low-risk — from 15.3% to 6.8%. This improvement suggests that the

[1] *** denotes $p<0.001$, indicating a high level of statistical significance.

system is more effective in preventing potential family crises, with the advantages significantly outweighing the administrative costs incurred by managing a relatively small number of first-type errors (false positives).

b. Improvement in Timeliness of Service Interventions

The implementation of a smart early warning system has reduced the average response time for emergency cases from 24 hours to 8 hours, while the success rate of emergency interventions has increased from 76% to 91%. This improvement is critical in preventing crises such as domestic violence and child abuse.

c. Optimization of Resource Allocation

The integration of geographic information systems has enhanced the efficiency of visit route planning for social workers by 38%, while also facilitating more accurate matching of local social welfare resources, resulting in a 42% increase in resource utilization efficiency.

(3) User Satisfaction Survey Results

A satisfaction survey conducted among 186 social workers yielded the following results:

Table 6

User Satisfaction Assessment

Evaluation Aspect	Average Score[2]	Standard Deviation	Satisfaction Rate (≥4 points)
System Usability	4.2	0.8	78.4%
Functional Practicality	4.5	0.7	84.2%
Prediction Accuracy	4.1	0.9	75.6%
Work Efficiency Improvement	4.6	0.6	89.3%
Professional Skill Enhancement	3.9	1.0	68.7%
Overall Satisfaction	4.3	0.8	81.5%

The survey results reveal that social workers rated their overall satisfaction with the smart decision-making platform at 4.3 points (out of 5), with 81.5% of respondents rating their satisfaction at 4 or above. The aspect of work efficiency improvement received the highest rating (4.6 points), with 89.3% of users expressing satisfaction. These findings are consistent with the quantitative analysis results, underscoring the significant impact of AI technology on enhancing work efficiency.

Conversely, satisfaction regarding the enhancement of professional capabilities was comparatively lower (3.9 points), with only 68.7% of users expressing satisfaction. In-depth interviews indicated that some social workers harbor concerns that an over-reliance on AI systems may undermine their professional judgment abilities, a matter that merits further investigation.

3. Analysis of Algorithm Fairness and Bias

(1) Testing for Predictive Differences Among Groups

This study examines the predictive fairness of the AI system across various social groups. Through statistical parity analysis, we assessed predictive differences across dimensions such as gender, age, ethnicity, and region.

Statistical analysis indicates that the AI system's risk predictions did not demonstrate statistically significant systemic bias across major social variables, including gender, age, ethnicity, and region (all p-values > 0.05). This suggests that the algorithms have effectively mitigated potential discriminatory factors during their design and training phases.

[2] Note: The scoring standard is a 5-point scale, with 5 indicating very satisfied and 1 indicating very dissatisfied.

Table 7

Analysis of Risk Prediction Differences Among Different Groups

Group Variable	Subgroup	High-Risk Prediction Rate (%)	Statistical Test	*p*-value	Conclusion
Gender	Male	22.4	χ^2=1.28	0.258	No significant difference
	Female	24.1			
Age	Youth (18-35)	26.8	χ^2=5.43	0.142	No significant difference
	Middle-aged (36-55)	21.9			
	Elderly (56+)	19.3			
Ethnicity	Hakka	21.2	χ^2=3.91	0.271	No significant difference
	Minnan	23.6			
	Mainland Chinese	25.1			
	Indigenous	28.4			
Region	Urban	22.7	χ^2=2.14	0.343	No significant difference
	Township	24.8			
	Remote	26.3			

However, it is noteworthy that, despite the absence of statistically significant differences, the high-risk prediction rate for the indigenous population (28.4%) remains higher than that of other groups, which may reflect genuine socioeconomic disadvantages rather than algorithmic bias. Future research should further delineate between "statistical differences" and "unfair discrimination."

(2) Calibration Analysis

Calibration serves as a critical metric for evaluating the fairness of AI systems, measuring the alignment between predicted probabilities and actual risk occurrence rates.

The calibration analysis indicates that the AI system maintained robust calibration between male and female groups, with the maximum calibration difference being merely 1.4 percentage points, significantly below the acceptable threshold of 5%. This finding suggests that the system's risk predictions possess comparable credibility across different gender groups.

Table 8

Calibration Analysis Among Different Groups

Predicted Risk Interval	Actual Incidence Rate - Male (%)	Actual Incidence Rate - Female (%)	Calibration Difference	Evaluation Result
0.0-0.1	3.2	3.8	0.6	Good Calibration
0.1-0.3	18.7	19.4	0.7	Good Calibration
0.3-0.5	38.6	37.9	-0.7	Good Calibration
0.5-0.7	59.2	58.1	-1.1	Good Calibration
0.7-0.9	78.4	79.8	1.4	Good Calibration
0.9-1.0	92.1	91.6	-0.5	Good Calibration

(3) Implementation Effects of Bias Mitigation Strategies

To further enhance algorithmic fairness, this study implemented several bias mitigation strategies:

a. **Data Level:** Resampling techniques were employed to balance the sample distribution among various groups, ensuring the representativeness of the training data.

b. **Algorithm Level:** Fairness constraints were incorporated into the loss function, utilizing adversarial debiasing

techniques.

c. **Post-Processing Level:** Decision thresholds for different groups were adjusted to ensure equalized opportunity.

Following the implementation of these strategies, the predictive differences among groups were further minimized, with the maximum difference decreasing from the original 6.2 percentage points to 3.7 percentage points, thereby demonstrating the effectiveness of the bias mitigation measures.

4. Professional Integration and Human-Machine Collaboration Model

(1) Decision Support Rather Than Decision Replacement

The findings from comprehensive interviews suggest that the efficacy of artificial intelligence (AI) applications is contingent upon the establishment of a human-machine collaboration framework that prioritizes "decision support" over "decision replacement." A significant majority, 83%, of the social workers interviewed contend that AI systems ought to furnish risk analyses and recommendations, while the ultimate decision-making authority should reside with professional social workers.

Interviewee A, a senior social worker with 15 years of experience, articulated, *"AI can swiftly analyze extensive datasets and highlight risk factors that I may overlook; however, each family's circumstances are distinct. The system lacks the capacity to comprehend intricate elements such as cultural context and family dynamics, which necessitate professional discernment."*

(2) Integration of Professional Knowledge and Data Insights

The research indicates that the optimal practice model merges the professional expertise of social workers with the data-driven insights provided by AI. This integration is particularly evident in the following stages:

a. **Initial Screening Stage:** The AI system performs extensive risk screenings to promptly identify potential high-risk cases.

b. **In-Depth Assessment Stage:** Social workers utilize AI recommendations in conjunction with their professional judgment to conduct thorough risk assessments.

c. **Intervention Planning Stage:** AI offers suggestions for resource matching, while social workers formulate individualized service plans tailored to the specific characteristics of each case.

d. **Follow-Up Assessment Stage:** AI tracks trends in risk alterations, whereas social workers assess the effectiveness of interventions and modify strategies accordingly.

(3) Capacity Building and Professional Development

The advent of AI technology has introduced both challenges and opportunities regarding the competencies required of social workers. Survey results reveal that 76% of social workers recognize the necessity to enhance digital literacy, and 68% advocate for improved data interpretation skills.

Table 9

Analysis of Social Workers' Capacity Building Needs

Competency Aspect	Importance Rating	Current Level	Gap	Priority Order
Data Interpretation Ability	4.3	2.8	1.5	1
System Operation Skills	4.1	3.2	0.9	3
AI Knowledge Understanding	3.9	2.5	1.4	2
Critical Thinking	4.6	3.8	0.8	4
Ethical Judgment Ability	4.7	4.1	0.6	5

In response to these identified needs, the Ministry of Health and Welfare has devised a series of capacity-building initiatives, which include digital literacy training, workshops on AI knowledge, and platforms for cross-disciplinary exchange (Chen et al., 2025b).

5. Ethical Challenges and Governance Framework

(1) Privacy Protection and Ethical Data Use

The operation of AI systems necessitates the collection of substantial amounts of personal sensitive data, thereby presenting significant challenges to privacy protection. This study identifies several critical ethical concerns:

a. **Complexity of Informed Consent:** Traditional models of informed consent are inadequate in addressing the multifaceted data usage methods employed by AI systems. Approximately 58% of service users report difficulty in comprehending how their data is utilized by AI systems.
b. **Data Minimization Principle:** The challenge remains to minimize the extent of data collection while ensuring predictive accuracy.
c. **Data Retention Period:** Questions regarding the appropriate duration for retaining personal data and the timing for its deletion remain inadequately addressed.
d. **Cross-Agency Data Sharing:** While inter-agency data sharing can enhance service quality, it simultaneously heightens the risk of privacy violations.

(2) Algorithm Transparency and Interpretability

Social workers express considerable concern regarding the opaque nature of AI systems, often referred to as the “black box.” Survey data indicates that 89% of social workers desire clarity on how AI systems derive specific conclusions, and 74% believe that the systems should provide explicit explanations.

To address this demand, this study has developed an interpretative interface based on SHAP values, which visually represents the contributions of various risk factors to the final predictive outcomes. User testing has demonstrated that this feature significantly enhances social workers’ trust and acceptance of the system.

To address this requirement, the present study incorporated XAI methodologies, specifically SHAP and LIME, during the analytical phase. Additionally, a visualization interface grounded in SHAP values was developed and integrated into the intelligent decision-making platform. This interface graphically delineates the influence of various risk factors — such as “history of domestic violence” and “unemployment status” — on the predictive outcomes for each case. User evaluations demonstrated that this feature substantially improved social workers’ trust in and acceptance of the system, facilitating their ability to comprehend and critically engage with AI-generated recommendations rather than adhering to them uncritically.

(3) Accountability and Accountability Mechanisms

In instances where AI systems generate erroneous predictions resulting in adverse outcomes, the issue of accountability becomes paramount. This study advocates for the establishment of a multi-tiered responsibility framework:

a. **Responsibility of Technology Developers:** Ensure the technical accuracy of algorithms and implement quality assurance mechanisms.
b. **Responsibility of System Operators:** Develop appropriate usage guidelines and provide comprehensive education and training.
c. **Responsibility of Professional Users:** Accurately comprehend the limitations of the system and base decisions on professional judgment.

d. Responsibility of Regulatory Authorities: Formulate suitable regulatory frameworks to ensure that the system adheres to ethical standards.

6. The Influence of Contextual, Procedural and Governance Dimensions on Trust in AI Systems

This research demonstrates that, notwithstanding the technical excellence of AI systems and the acquisition of users' informed consent, their overall usability and the degree of trust they engender are substantially shaped by three principal dimensions: contextual, procedural and governance factors.

a. Contextual Factors: Key contextual determinants include the extent to which organizational culture fosters innovation, the adequacy of managerial support, and the availability of peer support networks. These elements critically influence social workers' readiness to adopt the AI system.

b. Procedural Factors: The integration of the AI system into social workers' existing workflows must be seamless to avoid imposing additional burdens. For instance, functionalities such as automatic importation of case records and the generation of risk assessment reports with minimal user effort significantly enhance the system's operational practicality.

c. Governance Factors: Beyond securing informed consent, the establishment of a transparent governance framework is essential. This framework should encompass the formation of interdisciplinary ethics review committees, the implementation of clear accountability mechanisms—particularly in instances of erroneous predictions—and the provision of accessible channels for user complaints and feedback. Such governance structures are fundamental to transitioning from mere "passive consent" to a state of "active trust."

7. International Experience Comparison and Insights

(1) Experience from the United States: IBM Watson Care Manager

The Watson Care Manager system, developed by IBM in collaboration with the Child Welfare League in the United States, exemplifies the substantial potential of AI in child protection. The system employs natural language processing and machine learning algorithms to analyze information from various data sources, including hospitals, schools, and domestic violence centers, to identify potential cases of child abuse and neglect (Seniutis et al., 2024; Yang et al., 2022).

a. Key Success Factors

(a) Integration of diverse data sources
(b) Robust natural language processing capabilities
(c) Seamless integration with existing workflows
(d) Ongoing evaluation of effects and system optimization

b. Challenges and Limitations

(a) High initial investment costs
(b) Variability in data quality
(c) Complex coordination of inter-agency collaboration
(d) Ethical and privacy concerns

(2) Experience from the United Kingdom: PRIS

The Predictive Risk Intelligence System (PRIS) system developed in the UK is primarily utilized to predict the risk of child abuse by analyzing social service records, educational data, and health data to construct risk prediction models (Gawronski et al., 2024; Negri-Ribalta et al., 2024).

a. Innovative Features

(a) Utilization of randomized controlled trials to validate effectiveness
(b) Focus on algorithm fairness testing
(c) Establishment of a comprehensive ethical review mechanism
(d) Emphasis on a human-machine collaboration model

b. Implementation Results

(a) 27% increase in the accuracy of identifying high-risk cases
(b) 35% improvement in case handling efficiency
(c) 82% satisfaction rate among social workers

(3) Experience from Singapore: Integrated Social Service Platform

The integrated social service platform established by the Singapore government employs AI technology to optimize resource allocation and enhance service efficiency (Woo & Loo, 2024). The key features of the platform include:

a. Government-led unified platform
b. Integration of cross-departmental data
c. Focus on user experience design
d. Emphasis on cost-benefit analysis

(4) Insights for Taiwan

Drawing from international experiences, the following insights are proposed for the application of AI in social work in Taiwan:

a. Technical Development Strategy

(a) Implement a gradual development approach, initiating pilot projects in low-risk areas
(b) Prioritize system interoperability to prevent information silos
(c) Invest in infrastructure development to ensure data quality

b. Governance Mechanism Construction

(a) Establish mechanisms for cross-departmental coordination
(b) Formulate dedicated regulations for AI governance
(c) Create independent ethical review committees

c. Professional Capacity Enhancement

(a) Strengthen digital literacy training for social workers
(b) Develop mechanisms for cultivating cross-disciplinary professional talent
(c) Foster collaboration between academia and practice.

8. Technical Limitations and Directions for Improvement

(1) Limitations Pertaining to Data Quality and Availability

a. Insufficient Data Standardization: Variations in data formats and coding standards across different organizations hinder the effectiveness of system integration. It is advisable to implement unified data standards to facilitate the standardization of data formats.

b. Incomplete Historical Data: The absence or incompleteness of significant historical data adversely affects the efficacy of model training. It is recommended to enhance data collection mechanisms and establish processes for data quality verification.

c. Inadequate Timeliness: The infrequent updating of critical data compromises the timeliness of risk predictions. Establishing a real-time data update mechanism is suggested to improve data timeliness.

(2) Limitations Related to Algorithmic Technology

a. Addressing Long Tail Effects: The algorithm's capacity to manage rare yet high-risk scenarios is limited. It is proposed to integrate rule-based methods to improve the identification of anomalous situations.

b. Dynamic Risk Variability: Current models predominantly rely on static features and lack responsiveness to dynamic risk changes. The introduction of time series analysis techniques is recommended to develop dynamic risk prediction models.

c. Cross-Cultural Applicability: The effectiveness of algorithms across diverse cultural contexts requires further validation. Conducting cross-cultural comparative studies and adjusting model parameters accordingly is suggested.

(3) Limitations in System Integration and Scalability

a. Integration with Existing Systems: The integration of current social work information systems encounters both technical and institutional challenges. Adopting a microservices architecture is recommended to enhance system flexibility.

b. Scalability Considerations: As usage scales, challenges to system performance and stability arise. The adoption of cloud-native technologies to establish an automatic scaling mechanism is suggested.

c. Cross-Platform Compatibility: Compatibility issues across various devices and platforms negatively impact user experience. Employing responsive design is recommended to ensure consistency across platforms.

9. Challenges and Opportunities in Practical Application

(1) Organizational Change Management

The implementation of AI technology transcends mere technical considerations and necessitates a process of organizational change. This study identifies that successful AI applications require accompanying strategies for organizational change management.

a. Sources of Resistance to Change

(a) Fear and distrust of new technology (42% of respondents)
(b) Concerns regarding job displacement (35% of respondents)
(c) Inertia associated with existing work habits (38% of respondents)
(d) Conservatism within organizational culture (29% of respondents)

b. Facilitating Factors for Change

(a) Support from senior management (importance rating 4.6/5)
(b) Adequate education and training (importance rating 4.5/5)
(c) Peer support and experience sharing (importance rating 4.3/5)
(d) Clear demonstration of benefits (importance rating 4.4/5)

(2) Cost-Benefit Considerations

The implementation of AI systems necessitates a thorough consideration of cost-benefit dynamics. This study conducted a preliminary cost-benefit analysis, the results of which are presented in Table 10.

The findings indicate a benefit-cost ratio of 2.08 for the AI system, suggesting that the investment is economically viable. The primary sources of benefits include labor cost savings due to enhanced efficiency, social benefits stemming from improved service quality, and reduced crisis management costs attributable to risk prevention.

Table 10

Cost-Benefit Analysis of AI Systems (Annual)

Cost Item	Amount (NT$10,000)	Benefit Item	Amount (NT$10,000)
System Development	2,800	Labor Cost Savings from Efficiency Improvement	4,200
Hardware Equipment	1,200	Service Quality Improvement Benefits	2,800
Software Licensing	800	Cost Savings from Risk Prevention	3,600
Personnel Training	600	Resource Allocation Optimization Benefits	1,500
Maintenance and Operation	1,000	Other Indirect Benefits	1,200
Total Cost	**6,400**	**Total Benefits**	**13,300**
		Net Benefit	**6,900**
		Benefit-Cost Ratio	**2.08**

(3) Strategies for Sustainable Development

To ensure the sustainable application of AI technology in social work, this study proposes the following strategic recommendations:

a. Technological Sustainability

(a) Establish an open technology framework to prevent technological lock-in.

(b) Invest in autonomous technological capabilities to minimize external dependencies.

(c) Create mechanisms for technology assessment and updates.

b. Financial Sustainability

(a) Develop diversified funding sources.

(b) Implement a cost-sharing mechanism for operational expenses.

(c) Explore models for public-private partnerships.

c. Social Sustainability

(a) Ensure that technological advancements align with societal values.

(b) Uphold the roles and dignity of professional workers.

(c) Promote digital inclusion to mitigate the risk of digital divides.

IV. Conclusion

1. Summary of Research Findings

This investigation utilized a mixed-methods approach to examine the effects, challenges, and developmental prospects of machine learning technology in the risk classification of vulnerable populations. The findings reveal that the integration of artificial intelligence (AI) technology within the realm of social work yields substantial positive outcomes, albeit accompanied by significant ethical and practical challenges.

(1) Key Findings on Technical Effectiveness

The machine learning algorithms exhibited remarkable efficacy in risk prediction tasks, achieving an accuracy rate of 89.2% with deep neural network models, which represents a 17.2 percentage point enhancement over traditional manual assessment techniques. This notable advancement is critical for the early detection of high-risk cases and the prevention of crisis situations. An analysis of feature importance identified a history of domestic violence, unemployment status, and childcare responsibilities as the most significant predictors of risk,

corroborating existing social work theories and affirming that AI technology can effectively identify key risk factors in social work practice.

Time series analysis indicated that risk levels display considerable seasonal fluctuations, with the highest proportion of high-risk cases occurring in February (28.9%) and the lowest in May (18.6%). This trend provides valuable insights for resource allocation and workforce planning, facilitating the establishment of a more precise preventive service framework.

(2) Key Findings on Practical Effects

The implementation of an intelligent decision-making platform has profoundly transformed social work practice. The most notable impact is a marked increase in work efficiency, with case handling time reduced by 31.9%, risk assessment completion time decreased by 58.7%, and the average monthly case load for social workers rising by 35.7%. These enhancements primarily result from features such as automated risk assessments, intelligent home visit preparations, and speech-to-text capabilities.

Service quality improvements are also significant, as evidenced by a reduction in the missed diagnosis rate for high-risk cases from 15.3% to 6.8%, a decrease in emergency response time from 24 hours to 8 hours, and an increase in the success rate of emergency interventions from 76% to 91%. These advancements are vital for safeguarding vulnerable populations and averting crisis situations.

User satisfaction surveys reveal that 81.5% of social workers express satisfaction with the intelligent decision-making platform, particularly appreciating the enhancements in work efficiency (4.6/5, 89.3% satisfaction). However, satisfaction regarding the enhancement of professional capabilities is comparatively lower (3.9/5), reflecting social workers' apprehensions about the potential implications of AI technology on their professional development.

(3) Key Findings on Fairness and Ethics

An analysis of algorithmic fairness indicated that the AI system did not demonstrate statistically significant systemic bias across major social variables, including gender, age, ethnicity, and region. Calibration analysis further confirmed that the system maintains robust predictive reliability across diverse groups. Following the implementation of bias mitigation strategies, predictive disparities between groups were further minimized, illustrating the effectiveness of technical measures in fostering algorithmic fairness.

Nonetheless, the study also uncovered significant ethical challenges, including complexities surrounding privacy protection, inadequate algorithm transparency, and ambiguities in accountability. Notably, 58% of service users reported difficulties in comprehending how their data is utilized by the AI system, while 89% of social workers expressed a desire to understand the decision-making logic of the AI system, underscoring the necessity for enhanced system transparency and interpretability.

(4) Key Findings on Professional Integration

This study affirmed the significance of the "decision support rather than decision replacement" model of human-machine collaboration. A substantial 83% of social workers contend that the AI system should provide analyses and recommendations, while the ultimate decision-making authority should reside with professional social workers. A successful integration model necessitates the amalgamation of social workers' professional expertise with AI's data-driven insights, capitalizing on the strengths of both at various stages of practice.

The advent of AI technology has introduced new challenges regarding the competencies required of social workers, with 76% advocating for enhanced digital literacy and 68% emphasizing the need for improved data interpretation skills. This highlights the critical importance of capacity building and professional development in the context of AI application.

2. Theoretical Contributions and Innovations

(1) Theoretical Framework Construction

This study developed the "Theory Framework for the Intelligent Transformation of Social Work," which integrates technology acceptance theory, organizational change theory, and social work value theory, thereby providing a systematic theoretical foundation for comprehending the application of AI technology in social work. The framework underscores the balanced advancement of three dimensions: technical effectiveness, professional integration, and ethical considerations, establishing a significant theoretical basis for future research.

(2) Methodological Innovation

The mixed-methods approach employed in this study, which combines machine learning model evaluation, statistical analysis, qualitative interviews, and participatory observation, offers a methodological reference for research on AI applications in social work. Notably, the integration of algorithm fairness testing with social work ethics analysis provides valuable insights for interdisciplinary research.

(3) Theoretical Significance of Empirical Findings

The findings of this study enhance the understanding of the relationship between AI technology and the social work profession. Specifically, the validation of the "decision support rather than decision replacement" model and the exploration of the integration mechanism between professional judgment and data insights offer important theoretical guidance for the evolution of the social work profession in the digital era.

3. Practical Application Value

(1) Policy Formulation Guidelines

The outcomes of this study furnish empirical evidence for governmental formulation of AI governance policies. It is recommended to establish a multi-tiered responsibility division mechanism, a comprehensive ethical review process, and an adaptive regulatory framework to ensure the normative and sustainable application of AI technology.

(2) Institutional Implementation Guidelines

This study provides specific implementation guidelines for social work institutions regarding AI applications, including standards for technology selection, organizational change strategies, capacity-building programs, and effect evaluation mechanisms. These guidelines facilitate the smooth introduction of AI technology and the achievement of digital transformation within institutions.

(3) Professional Development Directions

This research delineates the developmental trajectory of the social work profession in the AI era, emphasizing the significance of professional judgment, critical thinking, ethical sensitivity, and digital literacy. This serves as a crucial reference for social work education and professional development.

4. Social Significance and Impact

(1) Enhancing the Effectiveness of Social Welfare Services

The application of AI technology markedly enhances the precision and timeliness of services for vulnerable populations, thereby strengthening the efficacy of the social safety net. The improvements in risk prediction accuracy and service efficiency will directly benefit numerous service users, fostering the realization of social equity and justice.

(2) Promoting Digital Inclusive Development

This study underscores the necessity of mitigating the digital divide and advancing digital inclusion. The proposed diverse training programs and accessible technology designs aim to ensure that all social groups can benefit from the advancements in AI technology.

(3) Advancing Smart Society Construction

This research provides significant insights for the establishment of a human-centered smart society, emphasizing the organic integration of technological advancement and humanistic care, ensuring that AI technology serves human welfare rather than merely pursuing technological progress.

5. Research Limitations and Future Research Directions

(1) Main Research Limitations

Nonetheless, this study exhibits certain limitations. Firstly, the representativeness of the sample may be constrained, as the underrepresentation of specific disadvantaged populations — such as individuals residing in remote regions or belonging to particular ethnic groups — could compromise the generalizability of the findings. Secondly, the assessment of long-term effects is insufficient; given that the intelligent decision-making platform has been operational for a relatively brief duration, it may not adequately capture its enduring impact. Furthermore, although the study highlights the fairness of the algorithm, there remains a need for more comprehensive investigation into methods for the ongoing monitoring and mitigation of algorithmic bias.

a. Limitations in Sample Representativeness: Although the study encompasses major social welfare institutions nationwide, certain special groups (e.g., those in remote areas or specific ethnic communities) remain underrepresented.

b. Limitations in Tracking Period: Given the relatively brief activation period of the intelligent decision-making platform, long-term effect evaluations necessitate additional time for verification.

c. Insufficient Cross-Cultural Comparisons: The research primarily focuses on the context of Taiwan, necessitating enhanced comparative analysis with other countries and regions.

d. Challenges of Rapid Technological Development: The swift evolution of AI technology may present timeliness challenges to the research findings due to ongoing technological updates.

(2) Future Research Recommendations

a. Long-Term Tracking Research: It is advisable to conduct longitudinal studies to assess the sustained effects and potential impacts of AI technology applications.

b. Cross-National Comparative Research: Future studies should engage in cross-national comparisons to explore the variances and commonalities of AI applications across diverse cultural and institutional contexts.

c. Emerging Technology Applications: Investigate the potential applications of emerging technologies such as generative AI, federated learning, and edge computing within social work.

d. User-Centered Research: Enhance user-centered research to gain deeper insights into the effects of AI technology on service recipients.

e. Ethical Framework Development: Continuously refine and enhance the ethical framework for AI applications, establishing a dynamic ethical assessment mechanism.

(3) Policy Recommendations and Practical Guidelines

a. Short-Term Recommendations (1-2 years)

(a) Enhance Regulatory Framework: Develop specific guidelines for AI applications in social work, clarifying regulations on critical issues such as data usage, privacy protection, and accountability.

(b) Strengthen Capacity Building: Implement systematic digital literacy training programs to bolster social

workers' capabilities in AI application.

(c) Establish Supervision Mechanisms: Create an AI application oversight committee to regularly evaluate system effectiveness and ethical compliance.

b. Mid-Term Recommendations (3-5 years)

(a) Promote Standardization: Establish a unified national data standard and technical specifications to facilitate system interoperability.

(b) Deepen Technology Applications: Explore the utilization of AI technology in additional areas of social work, such as community organization and policy analysis.

(c) Establish Evaluation Mechanisms: Develop a comprehensive evaluation framework for assessing the effects of AI applications to continuously optimize system performance.

c. Long-Term Recommendations (5 years and beyond)

(a) Build a Smart Ecosystem: Create a collaborative ecosystem involving government, academia, and industry to foster ongoing innovation.

(b) International Cooperation and Exchange: Strengthen international collaboration and engage in the development of a global AI governance framework, sharing Taiwan's experiences.

(c) Sustainable Development: Formulate a sustainable development model to ensure the long-term benefits of AI technology applications.

The intelligent transformation of social work practice signifies the onset of a new era. This study affirms that, when designed and implemented effectively, AI technology can substantially enhance the efficiency and effectiveness of social work, thereby providing improved services for vulnerable populations. However, this transformative process is accompanied by significant challenges, including algorithmic fairness, privacy protection, and professional integration.

Successful intelligent transformation necessitates a harmonious integration of technological innovation and humanistic care. AI technology should not supplant the core values and professional judgment inherent in social work but should serve as a powerful tool to augment professional capabilities and enhance service quality. Achieving this balance requires careful consideration of both technological advancement and ethical implications, fostering a synergistic relationship between efficiency enhancement and humanistic care.

Looking forward, the intelligent transformation of social work is poised to deepen. With the ongoing progression of technology and the accumulation of practical experience, there is reason to anticipate that AI technology will make substantial contributions to the establishment of a more equitable, efficient, and humane social welfare system. However, realizing this vision necessitates collaborative efforts from government, academia, practitioners, and the technology sector, alongside continuous research, innovation, and refinement.

This study serves as a pivotal starting point for these endeavors, yet numerous issues warrant further exploration. Future research is encouraged to continue advancing the theoretical framework and practical development of the intelligent transformation of social work, ultimately contributing to the creation of a more just society.

References

Altundağ, S. (2024). Artificial intelligence-based audit software: Today's realities and future vision. *Denetişim*, (31), 180–197.

Chen, D.K., Chang, H.C., & Chen, S.H. (2025a). Convergence of technological social changes in the development

of intelligent technology innovation system in Taiwan. *Journal of Scientometric Research*, *13*(3s), s66–s81.

Chen, K.M., Wang, S.T., Chao, S.R., & Kasirisir, K. (2025b). Link between social relationships and vulnerability among community-dwelling older adults living alone in Taiwan. *Asian Social Work and Policy Review*, *19*(1), e70000.

Crowley, R., Parkin, K., Rocheteau, E., Massou, E., Friedmann, Y., John, A., & Moore, A. (2025). Machine learning for prediction of childhood mental health problems in social care. *BJPsych Open*, *11*(3), e86.

Garcia-Lopez, A. (2024). Theory of mind in artificial intelligence applications. In *The Theory of Mind Under Scrutiny: Psychopathology, Neuroscience, Philosophy of Mind and Artificial Intelligence* (pp.723–750). Springer.

Gawronski, O., Briassoulis, G., El Ghannudi, Z., Ilia, S., Sánchez-Martín, M., Chiusolo, F., & Bestak, D. (2024). European survey on paediatric early warning systems, and other processes used to aid the recognition and response to children's deterioration on hospital wards. *Nursing in Critical Care*, *29*(6), 1643–1653.

Heeks, R., Wall, P.J., & Graham, M. (2025). Pragmatist-critical realism as a development studies research paradigm. *Development Studies Research*, *12*(1), 2439407.

Higgins, O., & Wilson, R.L. (2025). Integrating artificial intelligence (AI) with workforce solutions for sustainable care: A follow up to artificial intelligence and machine learning (ML) based decision support systems in mental health. *International Journal of Mental Health Nursing*, *34*(2), e70019.

Hsu, K., & Peng, L.P. (2023). Understanding vulnerability and sustainable livelihood factors from coastal residents in Taiwan. *Marine Policy*, *155*, 105793.

Negri-Ribalta, C., Lombard-Platet, M., & Salinesi, C. (2024). Understanding the GDPR from a requirements engineering perspective—a systematic mapping study on regulatory data protection requirements. *Requirements Engineering*, *29*(4), 523–549.

Rashid, A.B., & Kausik, A.K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*, 100277.

Seniutis, M., Gružauskas, V., Lileikiene, A., & Navickas, V. (2024). Conceptual framework for ethical artificial intelligence development in social services sector. *Human technology*, *20*(1), 6–24.

Tan, S., Taeihagh, A., & Baxter, K. (2022). The risks of machine learning systems. *arXiv preprint arXiv:2204.09852*.

Tsai, M.L., Chen, K.F., & Chen, P.C. (2025). Harnessing electronic health records and artificial intelligence for enhanced cardiovascular risk prediction: A comprehensive review. *Journal of the American Heart Association*, *14*(6), e036946.

Vasarhelyi, M.A., & Hsieh, S.F. (2020). The impact of artificial intelligence on the audit. *Computer Auditing, 42*, 59–68.

Woo, J.J., & Loo, D.R. (2024). *Building Urban Resilience: Singapore's Policy Response to Covid-19*. Taylor & Francis.

Yang, J., Chesbrough, H., & Hurmelinna-Laukkanen, P. (2022). How to appropriate value from general-purpose technology by applying open innovation. *California Management Review*, *64*(3), 24–48.