OPEN ACCESS
Middle East Research Journal of Economics and Management
ISSN 2789-7745 (Print) & ISSN 2958-2067 (Online)
Frequency: Bi-Monthly
DOI: https://doi.org/10.36348/merjem.2026.v06i04.002

KUWAIT SCHOLARS PUBLISHER
Website: http://www.kspublisher.com/
Email: office@kspublisher.com

An Explainable Multi-Index Financial Decision Framework for Personalized Household Financial Planning: Design and Simulation-Based Validation

Hendra Achmadi[1]*, Apriani Simatupang[1], Sylvia Samuel[1]
[2]Universitas Pelita Harapan

Abstract: Household financial planning requires simultaneous consideration of short-term liquidity, debt burden, long-term goals, risk-bearing capacity, and the prioritization of competing financial actions. Existing digital financial tools often treat these elements separately, while artificial-intelligence-based financial advisory systems introduce additional concerns regarding opacity, trust, and the traceability of recommendations. This study proposes the Explainable Multi-Index Financial Decision Framework (EMFDF), which integrates a Financial Health Index (FHI), Goal-Gap Index (GGI), Risk Capacity Index (RCI), and Financial Action Priority Score (FAPS) within a rule-constrained recommendation architecture. The framework is developed using a design science research orientation and evaluated through scenario-stratified synthetic household simulation. A total of 1,000 synthetic household profiles were generated to represent six decision contexts: emergency-fund inadequacy, excessive debt burden, insurance protection gaps, education-funding gaps, retirement-funding gaps, and long-term investment readiness. The deterministic recommendation engine achieved 92.7% agreement with the scenario benchmark. A random-forest surrogate model trained on the framework outputs achieved 89.0% holdout fidelity/accuracy, macro-precision of 0.888, macro-recall of 0.904, macro-F1 of 0.893, and macro-AUC of 0.988. Monotonicity tests confirmed that increases in emergency-fund adequacy did not reduce FHI, whereas increases in debt-service burden did not improve FHI. SHAP analysis identified the overall goal gap, debt-service ratio, education gap, emergency-fund adequacy, and retirement gap as the dominant drivers of model recommendations. The results demonstrate the internal coherence, transparency, and technical feasibility of the proposed framework as a proof-of-concept decision-support artifact. The study contributes a modular method for integrating household financial resilience, goal-based planning, risk capacity, constrained prioritization, and explainable machine learning. External validation using real households and professional financial planners is required before the framework can be interpreted as clinically or commercially validated financial advice.

Keywords: Explainable Artificial Intelligence, Household Finance, Financial Planning, Financial Resilience, Decision Support Systems, SHAP, Synthetic Data, Personalized Financial Advice.

Research Paper

***Corresponding Author:**
Hendra Achmadi
Universitas Pelita Harapan

How to cite this paper:
Achmadi, H., Simatupang, A., & Samuel, S. (2026). An Explainable Multi-Index Financial Decision Framework for Personalized Household Financial Planning: Design and Simulation-Based Validation. *Middle East Res J Econ Management, 6*(4), 105-118.
https://doi.org/10.36348/merjem.2026.v06i04.002

Article History:
| Submit: 11.08.2026 |
| Accepted: 14.09.2026 |
| Published: 17.09.2026 |

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

1. INTRODUCTION

Digital financial services increasingly mediate how individuals budget, save, borrow, invest, and plan for long-term financial goals. Recent research links digital financial literacy and access to digital financial services with financial well-being and financial resilience, indicating that technology can strengthen household financial capability when it is accompanied by appropriate knowledge and decision support (Kass-Hanna *et al.*, 2022; Mahdzan *et al.*, 2023; Choung *et al.*, 2023; Liu *et al.*, 2024). At the same time, the growth of algorithmic and AI-enabled financial advisory systems creates a new design problem: a system must not only produce a recommendation, but also demonstrate why the recommendation is financially coherent, appropriately constrained, and understandable to its user.

This requirement is particularly important in household financial planning. Unlike a narrowly defined credit-scoring or portfolio-selection task, household planning is multi-objective. A household may simultaneously face inadequate liquidity, high debt servicing, underinsurance, education obligations, retirement shortfalls, and opportunities for long-term investment. The preferred action therefore depends not on a single predictive target but on the interaction among current financial health, the distance between projected resources and future goals, the household's objective capacity to bear risk, and the urgency and affordability of each possible intervention. Financial resilience research similarly emphasizes the multidimensionality of household capacity to absorb and recover from adverse financial shocks (Liu & Chen, 2024; Peng & Liu, 2024; Liu *et al*., 2024).

Artificial intelligence can improve the scale and personalization of financial decision support, but the use of black-box models raises issues of interpretability and accountability. A recent systematic review of explainable artificial intelligence (XAI) in finance found that SHAP, feature-importance methods, and rule-based explanations are among the most frequently used approaches, while unresolved challenges remain regarding explanation quality, robustness, user trust, and the integration of explainability into domain-specific decision processes (Černevičienė & Kabašinskas, 2024). These concerns are reinforced by recent work on human-AI collaboration in finance, which identifies explainability, governance, calibrated trust, and human oversight as central research themes (Mirabile *et al*., 2026).

A second limitation in the current literature is the separation between technical prediction models and financial-planning logic. Purely data-driven models may produce statistically accurate outputs while failing to respect domain constraints, such as avoiding long-term investment recommendations when emergency reserves are critically inadequate or when debt-service burdens are excessive. Conversely, fully rule-based systems are transparent but may be rigid and difficult to personalize. Recent decision-support research therefore argues for socio-technical human-AI designs that combine domain knowledge, computational intelligence, and explicit evaluation of the artifact (Storey *et al*., 2024).

This study addresses these gaps by proposing the Explainable Multi-Index Financial Decision Framework (EMFDF) for personalized household financial planning. The framework integrates four components: (1) a Financial Health Index (FHI) for current financial condition; (2) a Goal-Gap Index (GGI) for the shortfall between projected resources and required goals; (3) a Risk Capacity Index (RCI) for objective risk-bearing capacity; and (4) a Financial Action Priority Score (FAPS) that ranks potential actions subject to hard financial-safety constraints. A machine-learning surrogate layer is then used to learn recommendation patterns, while SHAP provides global and local explanations of the features driving model outputs.

The study adopts a design science research orientation and simulation-based validation. Synthetic data are used deliberately as a proof-of-concept evaluation environment because the objective is to test logical coherence, sensitivity, classification behavior, and explainability before exposing real households to an unvalidated advisory system. This choice is consistent with the role of simulation in evaluating complex decision systems and with the broader use of synthetic data in financial AI when privacy, data scarcity, or controlled scenario coverage are relevant concerns. Importantly, the study does not claim external effectiveness or population-level prevalence from synthetic data; instead, it evaluates the internal validity and technical behavior of the proposed artifact.

The article addresses four research questions: RQ1, can current financial condition, future goal adequacy, and objective risk capacity be represented as interpretable household-level indices? RQ2, can these indices be integrated into a constrained prioritization algorithm that produces logically consistent personalized recommendations? RQ3, can a machine-learning model reproduce the recommendation structure with high predictive fidelity? RQ4, can SHAP-based explanations identify financially plausible drivers of the recommendations and improve auditability of the model?

2. Theoretical Background and Literature Review

2.1 Household Financial Well-Being and Financial Resilience

The theoretical foundation of the framework begins with contemporary financial well-being and financial resilience research. Financial well-being is not simply a function of income or wealth; it reflects the ability to manage current obligations, maintain financial security, pursue future goals, and reduce financial stress. Recent evidence shows that financial literacy, mental budgeting, self-control, and investment decision behavior contribute to financial well-being (Bai, 2023). Digital financial literacy likewise shows a positive association with financial well-being because digital capability affects individuals' ability to navigate financial services, interpret information, and protect themselves from digital risks (Choung *et al*., 2023).

Financial resilience extends this perspective by focusing on the household's capacity to withstand and recover from financial shocks. Liu and Chen (2024) conceptualize financial resilience through multiple dimensions rather than a single financial ratio. Liu *et al*., (2024) similarly show that financial literacy can strengthen household resilience through wealth and risk-mitigation channels. These studies justify a multidimensional view of financial condition. Accordingly, EMFDF does not equate financial health with a single savings or leverage ratio. Instead, the FHI combines liquidity, savings behavior, debt-service burden, cash-flow adequacy, and net-worth adequacy.

2.2 Goal-Based Household Planning

Household financial planning is inherently goal-based. Education, retirement, home ownership, insurance protection, and wealth accumulation have different horizons, required amounts, and funding trajectories. A household with an apparently healthy current cash flow may still be underfunded for retirement, while a household with substantial assets may be vulnerable because those assets are illiquid or heavily leveraged. Recent research on access to financial advice emphasizes the importance of planning behavior and comprehensive guidance, particularly for retirement preparedness (Chatterjee & Fan, 2023; Sunder *et al*., 2024).

The GGI operationalizes this temporal dimension by comparing the future value of resources allocated to a goal with the estimated future value of the required goal. The index is therefore not designed as a subjective perception score. It is a funding-adequacy measure intended to make the size of a shortfall explicit and comparable across goals after assumptions regarding inflation, return, horizon, and contribution rate have been specified.

2.3 Risk Tolerance versus Risk Capacity

Risk tolerance describes a person's willingness to accept uncertainty and potential loss; risk capacity refers to the household's objective ability to absorb loss without jeopardizing essential obligations or critical goals. The distinction is important because willingness to take risk should not override financial constraints. A household may report high tolerance for investment volatility but have low capacity because of weak liquidity, unstable income, high leverage, or a short planning horizon. EMFDF therefore treats risk tolerance as an informative profile variable but constructs the RCI from objective indicators such as investment horizon, emergency reserves, savings surplus, debt burden, income stability, and net-worth adequacy.

2.4 Explainable Artificial Intelligence and Calibrated Financial Advice

The use of AI in financial advisory changes the transparency requirement of decision support. In high-stakes or high-involvement domains, users and professional reviewers need to trace the link between input conditions and model outputs. Černevičienė and Kabašinskas (2024) show that explainability has become a major research stream in finance, with SHAP widely used to quantify feature contributions. However, explainability should not be treated merely as a visualization. In financial advice, an explanation should support auditability, reveal whether the model relies on financially plausible variables, and help detect situations in which a recommendation is driven by an inappropriate proxy.

Trust research further suggests that explanation quality should support calibrated reliance rather than unconditional trust. Han and Ko (2025) show in financial advisory experiments that trust changes dynamically with observed accuracy and that explanatory transparency can contribute to post-error trust repair. Mirabile *et al*., (2026) similarly identify explainability, AI governance, robo-advisory, and human oversight as interconnected themes in human-AI collaboration in finance. EMFDF therefore positions explanations as an accountability mechanism: the recommendation engine remains constrained by explicit financial rules, while SHAP is used to explain the behavior of the machine-learning layer rather than to justify unconstrained AI outputs after the fact.

2.5 Design Science Research Perspective

Design science research (DSR) is appropriate when the contribution is an artifact, method, model, or instantiation developed to solve an identified problem and evaluated for utility and performance. Contemporary decision-support scholarship emphasizes that DSR remains central to the design of human-AI systems because AI-based artifacts must balance technical performance, human interaction, and socio-technical constraints (Storey *et al*., 2024). The EMFDF is therefore treated as a method artifact comprising constructs (FHI, GGI, RCI, and FAPS), a decision model, a constrained recommendation procedure, and an explainability layer.

3. Proposed EMFDF Framework

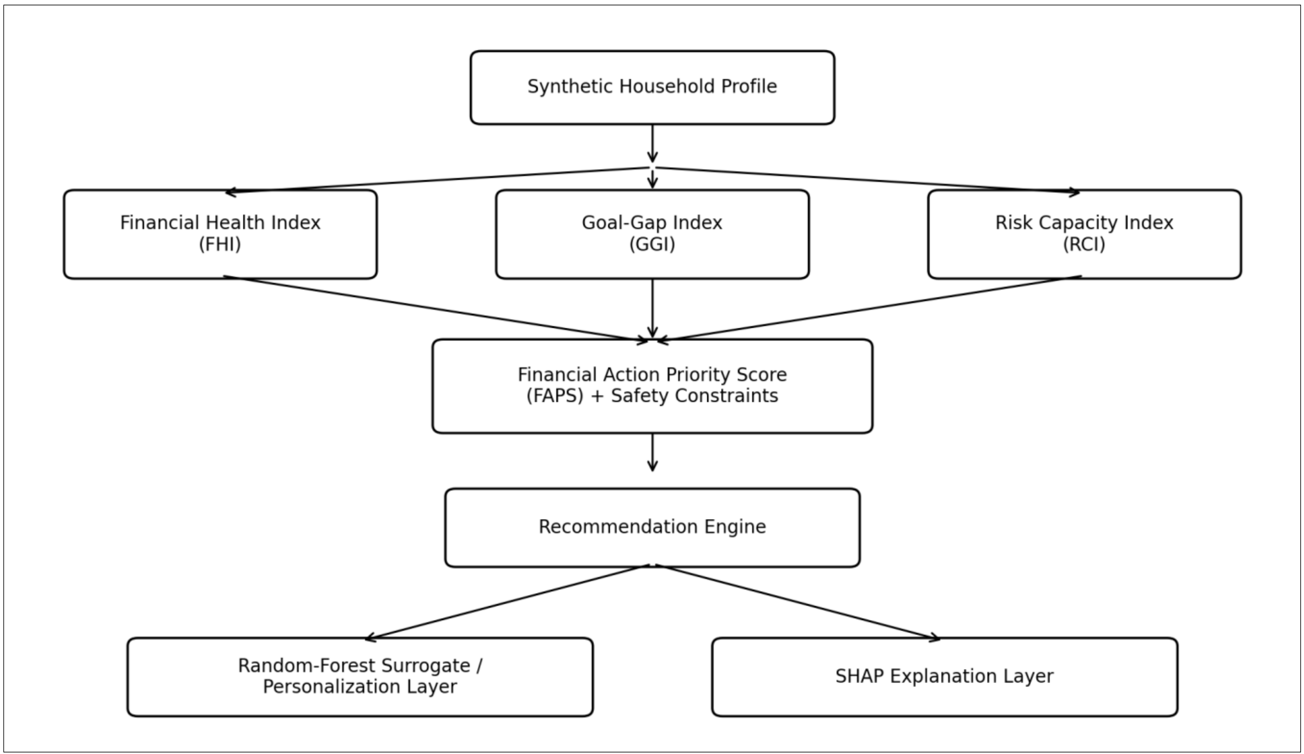

Figure 1: EMFDF multi-index decision and explainability architecture

3.1 Financial Health Index (FHI)

The FHI represents the household's current financial condition on a 0–100 scale. Five normalized components are used in the proof-of-concept implementation: liquidity adequacy (LQ), savings adequacy (SA), debt health (DH), cash-flow health (CFH), and net-worth adequacy (NWA). Each component is bounded to [0,100] to prevent extreme values from dominating the composite index.

$$FHI = 0.25LQ + 0.20SA + 0.20DH + 0.20CFH + 0.15NWA$$

Liquidity adequacy is based on emergency-fund months relative to a 12-month upper benchmark. Savings adequacy maps the savings rate into a bounded score. Debt health decreases as the debt-service ratio approaches a high-burden threshold. Cash-flow health reflects the available surplus after expenses and debt service. Net-worth adequacy compares household net worth with an income-scaled benchmark. The weights are design weights for the simulation and are not presented as universally optimal. Future field studies should recalibrate them through expert elicitation, analytic hierarchy process, best-worst method, or empirical estimation.

3.2 Goal-Gap Index (GGI)

For each financial goal j, the funding gap is calculated from the projected future value of current dedicated assets and planned contributions relative to the required future value of the goal.

$$GGI_j = \max[0,\ 1 - (FV(Current\ Assets_j) + FV(Contributions_j)) / FV(Required\ Goal_j)] \times 100$$

A score of 0 indicates that the goal is fully funded under the stated assumptions, whereas a score approaching 100 indicates a severe funding deficit. For multiple household goals, the proof-of-concept overall GGI uses the maximum of the education and retirement gaps to represent the most severe long-horizon shortfall. In a production implementation, a weighted aggregation could incorporate goal criticality, horizon, and user priority.

3.3 Risk Capacity Index (RCI)

The RCI captures objective financial risk-bearing capacity rather than psychological risk tolerance. The implementation combines investment horizon (H), emergency reserve adequacy (ER), savings surplus (SS), debt capacity (DC), income stability (IS), and net-worth adequacy (NWA).

$$RCI = 0.25H + 0.20ER + 0.20SS + 0.15DC + 0.10IS + 0.10NWA$$

The index is bounded to 0–100. A higher RCI indicates a stronger capacity to tolerate financial volatility without threatening essential commitments. The use of objective capacity constraints is particularly important for personalized systems because it prevents a high self-reported appetite for risk from becoming the sole basis for investment recommendations.

3.4 Financial Action Priority Score (FAPS)

EMFDF does not directly map each index into a product recommendation. Instead, it generates a priority score for each action category. The generic FAPS structure combines urgency (U), financial gap severity (G), expected financial impact (I), affordability or actionability (A), and risk penalty (R).

$$FAPS_k = \alpha U_k + \beta G_k + \gamma I_k + \delta A_k - \lambda R_k$$

The action k with the highest valid priority score becomes the top recommendation, subject to hard constraints. In the present simulation, six actions are evaluated: build emergency reserves, reduce debt, strengthen insurance protection, fund education, fund retirement, and increase long-term investment. Hard constraints reduce the long-term investment score when emergency reserves are below three months or when the debt-service ratio exceeds 40%. Education recommendations are disabled for households without dependents. These constraints operationalize the principle that a recommendation must be both individually relevant and financially admissible.

3.4.1 Formula Justification and Supporting Literature

Financial Health Index (FHI)

The FHI formula is specified as a composite score because household financial health is multi-dimensional rather than reducible to income alone. The selected components—liquidity adequacy, savings adequacy, debt health, cash-flow health, and net-worth adequacy—are supported by recent household-finance studies that emphasize liquid buffers, debt-service pressure, savings behaviour, and resilience under shocks. The weighting structure therefore operationalizes current solvency, short-term resilience, and long-term balance-sheet strength in one transparent index (Calcagnini *et al*., 2024; Carton *et al*., 2024; Liu *et al*., 2024).

Goal-Gap Index (GGI)

The GGI formula follows goal-based planning logic: a household is evaluated against the future value of required goals and the projected future value of dedicated assets and contributions. This design is consistent with goal-based investing and retirement-planning literature, where the relevant risk is not only return volatility but also the probability of failing to meet a specified financial goal at a specified time horizon (Kim *et al*., 2022; Simonato, 2023; Sinha & Irala, 2025).

Risk Capacity Index (RCI)

The RCI is designed to measure objective risk-bearing capacity, not psychological risk tolerance. Investment horizon, emergency reserves, savings surplus, debt capacity, income stability, and net-worth adequacy are used because financial-planning decisions should account for whether the household can absorb volatility, liquidity shocks, and income disruption before increasing investment risk (Calcagnini *et al*., 2024; Chatterjee & Fan, 2023; Kim *et al*., 2022).

Financial Action Priority Score (FAPS)

The FAPS algorithm is formulated as a constrained priority-ranking mechanism rather than an unconstrained product recommender. Urgency, goal gap, impact, affordability, and action risk are combined so that the recommendation is auditable and explainable. This is aligned with recent XAI-in-finance research emphasizing transparent decision logic, feature contribution, and human-auditable recommendations in high-stakes financial domains (Černevičienė & Kabašinskas, 2024; Sabharwal *et al*., 2025; Weber *et al*., 2024).

3.5 Explainability Layer

A random-forest model is trained as a data-driven surrogate of the constrained recommendation engine. The purpose is not to replace the deterministic financial logic in the proof-of-concept, but to test whether recommendation patterns can be learned from multidimensional profiles and to create a platform for future personalization. SHAP values are then calculated to quantify the direction and magnitude of each feature's contribution to model predictions. Global SHAP importance is used to audit which variables dominate the model overall, while local SHAP values can be used in the application to explain individual recommendations.

4. METHODOLOGY

4.1 Research Design

The study follows a DSR-oriented build-and-evaluate logic. The build phase defines the four financial constructs, their computational relationships, safety constraints, and recommendation architecture. The evaluation phase uses controlled simulation to examine internal coherence, sensitivity, recommendation agreement, predictive fidelity, and explainability. The study is therefore a simulation-based methodological validation rather than a clinical trial, consumer experiment, or population survey.

4.2 Synthetic Household Data Generation

A total of 1,000 synthetic household profiles were generated using a fixed random seed to support reproducibility. Variables included age, monthly income, monthly expenses, debt-service ratio, emergency-fund months, savings rate, net worth, number of dependents, income stability, financial horizon, risk tolerance, retirement-funding gap, and education-funding gap. The profiles were scenario-stratified rather than drawn as a simple random population sample. This was done to ensure sufficient representation of all six decision contexts and to prevent rare but important financial conditions from disappearing in the validation dataset.

Variable	Operational range / role
Age	25–60 years
Monthly income	IDR 6–100 million, bounded log-normal generation
Debt-service ratio	0–60% of monthly income
Emergency reserve	0–18 months
Savings rate	approximately −15% to 45%
Net worth	IDR 20 million–8 billion
Dependents	0–5
Income stability	0–1 normalized score
Planning horizon	approximately 3–30 years
Retirement gap	0–100% underfunding
Education gap	0–100% underfunding where applicable

The six scenario benchmarks were defined ex ante: (1) emergency-fund inadequacy; (2) excessive debt burden; (3) insurance protection need; (4) education-funding shortfall; (5) retirement-funding shortfall; and (6) long-term investment readiness. Profiles within each benchmark were generated from parameter regions consistent with the underlying scenario. This construction produces a controlled challenge set. It does not imply that the synthetic class proportions correspond to real-world population prevalence.

4.3 Validation Tests

Six validation procedures were used. First, monotonicity tests assessed whether directionally favorable changes produced non-adverse index changes. Emergency-fund adequacy was increased by three months while other components were held constant, and debt-service burden was increased by ten percentage points. Second, the FAPS recommendation engine was compared with the ex ante scenario benchmark. Third, a Random-Forest classifier with 500 trees, balanced class weighting, and a minimum leaf size of two was trained on 80% of the simulated observations and evaluated on a stratified 20% holdout sample. Fourth, model performance was assessed using holdout fidelity/accuracy, macro-precision, macro-recall, macro-F1, one-vs-rest macro-AUC, and a confusion matrix. Fifth, stratified five-fold cross-validation was used as a robustness check. Sixth, SHAP TreeExplainer was applied to the holdout data to estimate global and local feature contributions.

4.4 Predictive Performance and Explainability Evaluation Protocol

The Random-Forest layer is evaluated as a surrogate and personalization model that learns a mapping from a household profile to the financial action selected by the constrained EMFDF/FAPS policy. It does not predict FHI, GGI, RCI, or FAPS because those quantities are already computed deterministically by the framework. Instead, the dependent variable is the recommended financial action. Six classes are defined: (0) Build Emergency Fund, (1) Reduce Debt, (2) Increase Education Funding, (3) Increase Retirement Funding, (4) Increase Investment, and (5) Insurance/Protection Review. This separation makes the analytical sequence explicit: the EMFDF indices summarize household conditions; FAPS and hard constraints generate the benchmark policy action; the Random-Forest learns that recommendation structure; and SHAP subsequently explains the learned prediction.

Table 1: Recommendation classes used as the Random-Forest target variable

Code	Recommendation class	Interpretation
0	Build Emergency Fund	Strengthen short-term liquidity before higher-risk or longer-horizon actions.
1	Reduce Debt	Lower excessive debt-service burden and improve cash-flow resilience.
2	Increase Education Funding	Address a material education-goal funding gap.
3	Increase Retirement Funding	Address a material retirement-goal funding gap.
4	Increase Investment	Allocate investable surplus after safety constraints are satisfied.
5	Insurance/Protection Review	Address material protection gaps for the household.

The 1,000 scenario-stratified synthetic household profiles were divided using an 80% training and 20% testing split. Stratification preserved the representation of all six recommendation classes in both subsets, yielding 800 training cases and 200 holdout cases. The model uses raw household variables such as age, income, expenses, debt-service ratio, emergency-fund months, savings rate, dependents, and protection gap, together with the derived FHI, GGI, and RCI indices. The holdout sample is never used to fit the model. In addition, five-fold stratified cross-validation is used to assess whether performance is stable across alternative training partitions.

Model performance is evaluated with complementary multiclass metrics. Holdout accuracy equals the number of holdout observations for which the Random-Forest prediction matches the EMFDF/FAPS recommendation divided by the number of holdout observations. Precision is TP/(TP+FP), recall is

TP/(TP+FN), and F1 is 2(Precision × Recall)/(Precision + Recall). Because the problem contains six classes, macro-precision, macro-recall, and macro-F1 are reported so that each type of financial action receives equal weight. One-vs-rest macro-AUC measures discrimination across the six action classes. The confusion matrix reveals which actions are most frequently confused and therefore provides a substantive diagnostic of the policy-emulation task.

The central interpretation of predictive performance is surrogate fidelity rather than real-world advisory accuracy. Fidelity is defined as: $\text{Fidelity} = \Sigma\, I[\hat{y}_i = y_i^{EMFDF}] / N$, where $\hat{y}_i$ is the Random-Forest prediction and y_i^{EMFDF} is the deterministic EMFDF/FAPS policy output. A high fidelity score indicates that the learned model reproduces the designed recommendation policy. It does not establish that the recommendation is objectively optimal for real households, because the benchmark is generated by the framework rather than by independent professional-advisor judgments or observed financial outcomes.

SHAP is applied after prediction to answer a different question from the Random-Forest classifier. The classifier answers “Which financial action is recommended?”, whereas SHAP answers “Why did the model assign this household to that recommendation?” Global SHAP analysis aggregates the absolute contribution of each feature across all holdout observations and therefore identifies the variables that most consistently drive model behavior. Local SHAP analysis decomposes a single prediction into feature-level contributions, showing which values increased or reduced the probability of a particular action. The explanation layer therefore connects a personalized recommendation to measurable household conditions and supports traceability, auditability, and calibrated reliance (Černevičienė & Kabašinskas, 2024; Weber *et al*., 2024; Han & Ko, 2025).

An additional model-comparison and ablation experiment is used to assess the incremental value of the multi-index representation. Model A uses raw household variables only, whereas Model B adds FHI, the GGI measures, and RCI. Further ablation runs remove FHI, the GGI family, or RCI from Model B one at a time. A decline in macro-F1 after removing an index indicates that the corresponding representation contributes information that helps the surrogate emulate the EMFDF policy. This analysis must nevertheless be interpreted carefully: because FAPS is built from financial concepts that overlap with FHI, GGI, and RCI, strong performance is expected. The tests demonstrate policy emulation and representation value, not independent forecasting validity. This distinction is necessary to avoid circularity or data-leakage claims.

4.5 Reproducibility and Scope of Inference

All reported numerical results in this manuscript were produced from the simulation described above; they are not hypothetical values. Nevertheless, their scope of inference is intentionally narrow. They indicate whether the proposed artifact behaves consistently in a designed synthetic environment. They do not establish consumer benefit, financial outcome improvement, professional suitability, regulatory compliance, or superiority to human financial planners. Those claims require subsequent validation using real household data, expert review, and prospective user studies.

5. RESULTS

5.1 Index Distributions

Across the 1,000 synthetic profiles, mean FHI was 56.53 (SD = 20.02), mean GGI was 50.42 (SD = 28.44), and mean RCI was 56.87 (SD = 16.68). The ranges were broad by design: FHI ranged from 14.04 to 99.51, GGI from 0.91 to 99.94, and RCI from 20.07 to 97.06. This spread indicates that the scenario-generation procedure successfully created financially heterogeneous profiles rather than a narrow cluster around average conditions.

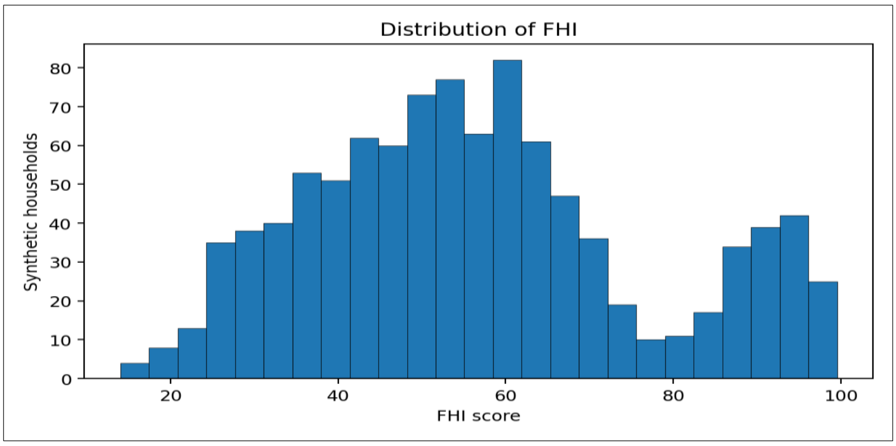

Figure 2: Distribution of Financial Health Index scores in the synthetic validation dataset

5.2 Monotonicity and Internal Consistency

The monotonicity tests produced the expected directional behavior in 100% of simulated cases. Adding three months of emergency reserves never reduced the FHI, while increasing the debt-service ratio by ten percentage points never increased the FHI. These tests do not prove that the index weights are optimal, but they provide evidence that the implemented transformations satisfy two fundamental financial logic constraints.

5.3 Recommendation-Engine Agreement

The deterministic FAPS recommendation engine achieved 92.7% agreement with the scenario benchmark. The strongest alignment occurred for education-funding cases (166 of 167), debt-reduction cases (158 of 167), long-term investment cases (157 of 166), and retirement-funding cases (146 of 166). Most disagreements involved the insurance-protection category, which sometimes outranked emergency-fund or retirement actions in profiles with multiple simultaneous vulnerabilities. This result is substantively informative: household financial planning is a multi-need prioritization problem, and overlapping vulnerabilities create legitimate boundary cases rather than simple labeling errors.

Benchmark scenario	Correct top recommendation	Approx. benchmark cases	Interpretation
Emergency Fund	134	167	Some cases shifted to insurance protection when dependents and income instability were high.
Debt Reduction	158	167	High debt burden strongly activated the debt-priority rule.
Insurance Protection	166	167	Protection need dominated in the targeted dependency/instability scenarios.
Education Funding	166	167	Large education gaps produced highly stable prioritization.
Retirement Funding	146	166	Some cases shifted to insurance protection under overlapping vulnerabilities.
Long-term Investment	157	166	Safety constraints prevented investment from dominating vulnerable profiles.

5.4 Machine-Learning Fidelity

The Random-Forest surrogate was trained on 800 profiles and evaluated on 200 stratified holdout profiles. It achieved 89.0% holdout fidelity/accuracy, macro-precision of 0.888, macro-recall of 0.904, macro-F1 of 0.893, and one-vs-rest macro-AUC of 0.988. Five-fold stratified cross-validation produced a mean macro-F1 of 0.871 (SD = 0.008). These results indicate that the recommendation structure can be learned with relatively high fidelity, although the performance is explicitly fidelity to the EMFDF/FAPS policy rather than external accuracy against independent financial-planner decisions.

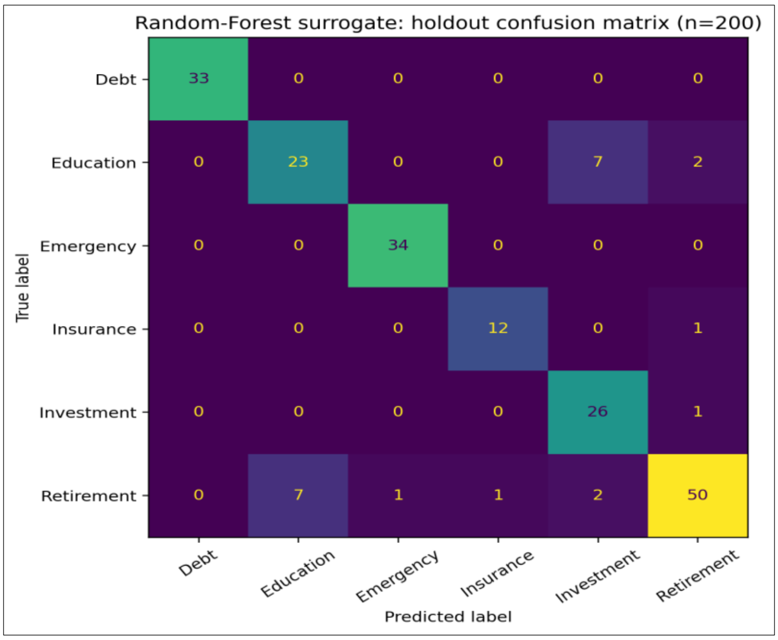

Figure 3: Random-Forest surrogate confusion matrix for the 20% holdout sample

The confusion matrix shows that Emergency Fund and Debt Reduction were reproduced almost perfectly in the holdout sample, whereas the largest overlaps occurred between Education Funding and Investment Allocation and between Retirement Investment and Education Funding. This pattern is substantively plausible because households may simultaneously face long-horizon funding gaps and investable-surplus decisions. In contrast, frequent confusion between Emergency Fund and Investment Allocation would signal a more serious violation of the framework's liquidity-first safety logic; such a pattern was not dominant in the present holdout results.

5.4.1 Incremental Value of the Multi-Index Representation

To examine whether the derived indices provide incremental representation value, the Random-Forest was re-estimated under a raw-variable specification and several ablation specifications. Model A uses only raw household variables. Model B adds FHI, the GGI measures, and RCI. The ablation models remove one index family at a time. The comparison is not a claim of causal importance; it evaluates whether the indices make the deterministic policy easier for the surrogate to reproduce.

Model A achieved macro-F1 = 0.840, whereas Model B achieved macro-F1 = 0.893, an increase of 0.054. Removing the GGI family produced the largest deterioration (macro-F1 = 0.820), followed by removing RCI (0.874) and FHI (0.885). Within the synthetic policy environment, these results suggest that the multi-index representation makes the recommendation structure more learnable, with goal-gap information contributing the largest incremental signal. Because the target is produced by EMFDF itself, the result should be described as representation/fidelity evidence rather than independent predictive validation.

Table 2: Model comparison and ablation analysis

Specification	Holdout accuracy	Macro-F1
Model A: raw household variables only	0.835	0.840
Model B: raw variables + FHI + GGI + RCI	0.890	0.893
Ablation: Model B without FHI	0.880	0.885
Ablation: Model B without GGI family	0.810	0.820
Ablation: Model B without RCI	0.870	0.874

Metric	Result
Recommendation-engine agreement with scenario benchmark	92.7%
Random-Forest holdout fidelity/accuracy (80/20 split)	89.0%
Macro-precision	0.888
Macro-recall	0.904
Macro-F1	0.893
Macro-AUC (OvR)	0.988
Five-fold CV macro-F1	0.871 ± 0.008
FHI monotonicity: + emergency reserve	100% directionally consistent
FHI monotonicity: + debt burden	100% directionally consistent

Table 3: Interpretation of the predictive-performance measures used for the Random-Forest surrogate

Measure	What it answers in EMFDF
Accuracy	How often does the surrogate reproduce the benchmark recommendation on the holdout sample?
Macro-precision	When each action is predicted, how often is that prediction correct, giving equal weight to all classes?
Macro-recall	How completely does the model recover each benchmark action, giving equal weight to all classes?
Macro-F1	How well does the model balance precision and recall across recommendation classes?
Macro-AUC	How well does the model discriminate each action from all other actions across probability thresholds?
Confusion matrix	Which financial actions are most frequently mistaken for one another?
Surrogate fidelity	How closely does the learned model emulate the deterministic EMFDF/FAPS policy?

5.5 SHAP Explanation Results

Global SHAP analysis identified emergency-fund months as the strongest overall driver, followed by debt-service ratio, insurance gap, education GGI, FHI, overall GGI, retirement GGI, and RCI. The ranking is financially plausible because liquidity adequacy, leverage, protection needs, and goal underfunding directly determine the urgency of household financial actions. The result also indicates that the surrogate is not driven primarily by simple demographic segmentation. Instead, financial-state and goal-adequacy variables dominate recommendation behavior.

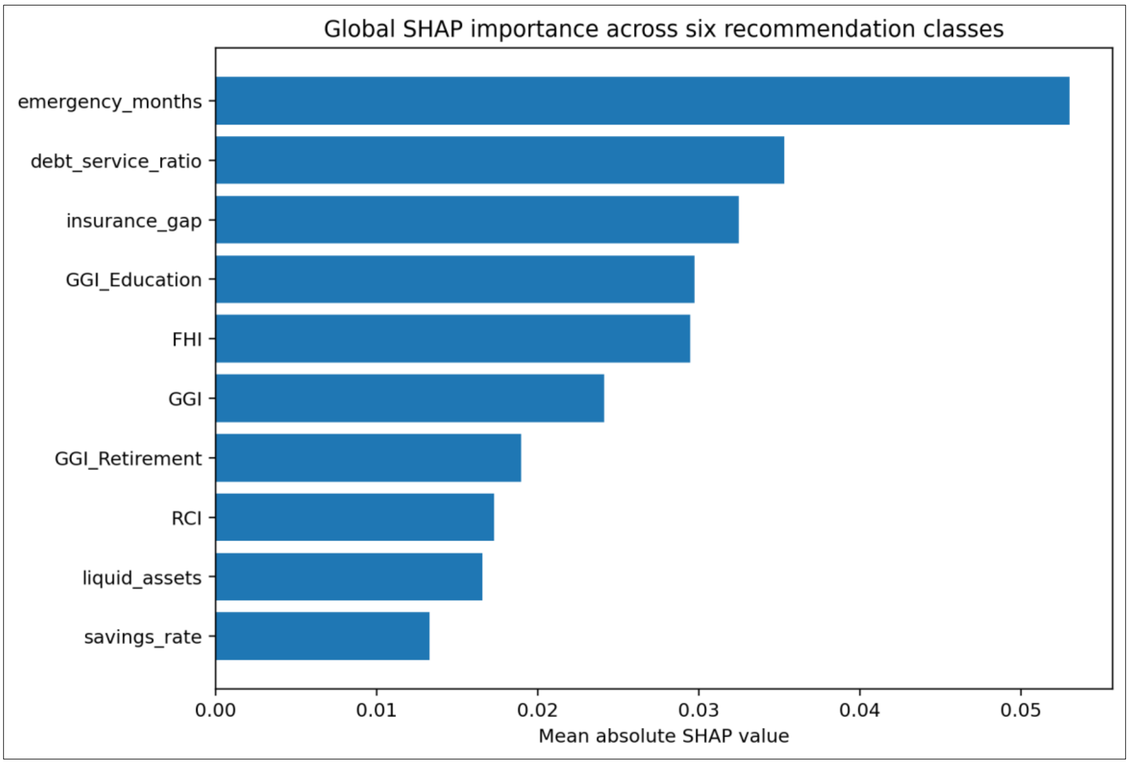

Figure 4: Global SHAP feature importance based on mean absolute SHAP values across the holdout sample and all six classes

5.6 Interpreting Predictive Performance: A Worked Example

The predictive and explanation results are easier to interpret when linked to a concrete holdout observation. The following example uses synthetic household Case 613 from the 20% test sample; it is therefore an actual observation used in the reported holdout evaluation rather than a hypothetical example. The deterministic EMFDF/FAPS policy and the Random-Forest surrogate both selected "Build Emergency Fund."

Table 4: Observed holdout household profile used to explain prediction and SHAP output

Variable	**Observed holdout value**
Case ID	613
Age	42 years
Monthly income	IDR 64,000,000
Monthly expense	IDR 34,836,525
Debt-service ratio	10.2%
Emergency-fund coverage	0.49 months
Savings rate	35.4%
FHI	62.92
GGI	33.60
RCI	59.00
Dependents	0

For this case, the constrained FAPS engine selected "Build Emergency Fund" as the benchmark policy action. The Random-Forest received the same profile and produced P(Emergency Fund) = 0.984, P(Retirement Investment) = 0.008, P(Investment Allocation) = 0.003, with the remaining classes jointly accounting for less than 0.005. The predicted class was therefore "Build Emergency Fund," matching the EMFDF/FAPS policy output. This observation is counted as a correct surrogate prediction because the learned model reproduces the deterministic policy decision.

In the full holdout sample, 178 of 200 cases were reproduced correctly, giving fidelity/accuracy of 89.0%. Accuracy alone is insufficient in a six-class problem; therefore macro-precision (0.888), macro-recall (0.904), macro-F1 (0.893), and macro-AUC (0.988) were also evaluated. Macro averaging gives each recommendation class equal importance, preventing a more frequent class from dominating the overall evaluation.

The confusion matrix provides a substantive diagnostic rather than merely a technical score. Confusion between education and retirement funding, for example, can indicate that both needs are simultaneously severe in some households. By contrast, repeated confusion between emergency-fund building and aggressive investment allocation would indicate a more serious policy-emulation problem because the two actions reflect different financial-safety priorities.

5.7 Worked Example of a Local SHAP Explanation

SHAP decomposes the prediction for synthetic Case 613 into feature-level contributions for the "Build Emergency Fund" class. Positive SHAP values increase the class output relative to the model baseline, while negative values reduce it. Because this case comes from the actual holdout analysis, the values below are empirical model explanations rather than illustrative placeholders.

Table 5: Local SHAP explanation for the observed emergency-fund recommendation

Feature	Observed SHAP contribution	Interpretation
emergency_months	+0.413	The very low reserve coverage strongly increases the emergency-fund recommendation.
liquid_assets	+0.161	Low liquid assets reinforce the immediate liquidity shortfall.
insurance_gap	+0.056	Protection needs contribute, but do not override the liquidity-first constraint.
FHI	+0.037	The overall financial-health state contributes to the priority assigned by the surrogate.
debt_service_ratio	+0.033	Debt-service pressure contributes to short-term financial vulnerability.
RCI	+0.026	Objective risk-bearing capacity modifies the relative urgency of the emergency-fund action.

The local explanation can be translated into an auditable narrative: "For synthetic Case 613, the recommendation to build an emergency fund is driven primarily by extremely low emergency-fund coverage (0.49 months) and limited liquid assets, with additional contributions from protection needs, financial-health status, and debt-service burden. The explanation therefore links the recommendation to measurable conditions rather than presenting an unsupported AI suggestion." This example illustrates how the SHAP layer can provide a household-specific reason for a recommendation while the deterministic EMFDF rules remain the underlying safety boundary.

At the global level, mean absolute SHAP values are aggregated across all holdout observations. A global ranking can therefore answer a different question: which variables most consistently drive recommendation behavior across households? Local and global explanations should be reported together because a variable that is influential on average may not be the main reason for a particular household recommendation.

6. DISCUSSION

6.1 Interpretation of the Multi-Index Design

The simulation results support the internal coherence of the proposed multi-index design. The FHI captures present financial condition, the GGI captures future underfunding, and the RCI captures risk-bearing capacity. These constructs serve different functions and should not be collapsed into one score. A financially healthy household can still have a large retirement gap; a household with a strong risk appetite can still have weak risk capacity; and a household with substantial income can still be financially vulnerable if debt servicing and liquidity are poor. The architecture therefore reflects recent household-finance research emphasizing multidimensional financial well-being and resilience (Liu & Chen, 2024; Liu *et al.*, 2024; Carton *et al*., 2024).

6.2 Why Constrained Recommendation is Preferable to Unconstrained AI Advice

A central design contribution is the separation between financial computation, priority logic, and machine-learning explanation. In many generative-AI implementations, the language model is implicitly asked to perform all three functions. Such a design is difficult to audit and may produce recommendations that are fluent but financially inconsistent. EMFDF instead treats deterministic calculators and safety constraints as the authoritative layer, with the machine-learning model used to learn patterns and the explanation layer used to reveal those patterns. This architecture is compatible with the broader direction of human-AI decision-support research, which favors explicit allocation of decision roles and accountability rather than indiscriminate automation (Storey *et al*., 2024; Mirabile *et al.,* 2026).

6.3 Explainability as Financial Auditability

The SHAP results are meaningful not because SHAP itself is novel, but because explainability is embedded into a domain-constrained financial decision process. Recent XAI research in finance shows that feature attribution is widely used, but also stresses the need for methods to be matched to the decision task and for explanations to be evaluated in context (Černevičienė & Kabašinskas, 2024). In EMFDF, SHAP can serve three audiences. For users, it can support an understandable answer to the question 'why is this my priority?'. For financial planners, it can provide an audit trail identifying the model's dominant drivers. For developers and regulators, it can help identify undesirable dependence on unstable or inappropriate variables.

6.4 Role of Synthetic Validation

Simulation is particularly useful at an early design stage because it makes boundary conditions observable. Real household datasets often contain imbalanced conditions, missing values, privacy-sensitive fields, and limited examples of extreme combinations. Scenario-stratified synthetic data allow the designer to create controlled combinations of high debt, low liquidity, severe goal gaps, high dependency, and investment readiness. However, synthetic validation has a strict limitation: success against designed scenarios can only establish internal consistency. It cannot determine whether the chosen thresholds, weights, or recommendations are socially optimal or professionally accepted. Accordingly, the present study should be positioned as proof-of-concept validation prior to external testing.

7. Theoretical and Practical Contributions

7.1 Theoretical Contribution

The first theoretical contribution is the integration of three distinct financial states—current health, future goal adequacy, and objective risk capacity—into a single decision-support framework without reducing them to one latent construct. This clarifies that financial health, goal readiness, and risk capacity represent complementary dimensions of household financial decision-making.

Second, the study connects household financial resilience with explainable AI. Existing household-finance studies emphasize literacy, behavior, resilience, and access to advice, whereas XAI studies in finance predominantly focus on credit risk, fraud, stock prediction, and related institutional tasks (Černevičienė & Kabašinskas, 2024). EMFDF shifts the explainability problem to household planning and asks how a recommendation can be both personalized and constrained by explicit financial logic.

Third, the study conceptualizes explainability as part of the artifact rather than as a post hoc communication feature. The SHAP layer is linked to a financial action-priority algorithm whose admissibility constraints are visible. This design supports calibrated trust, consistent with recent research showing that transparency and accuracy jointly shape trust in AI-enabled financial advice (Han & Ko, 2025; Mirabile *et al*., 2026).

7.2 Methodological Contribution

Methodologically, the study demonstrates a staged validation procedure for financial decision artifacts: index monotonicity, scenario agreement, machine-learning fidelity, and explanation audit. This sequence can be reproduced before field deployment and can help identify logical defects prior to collecting costly or privacy-sensitive user data. The scenario-stratified approach also makes class coverage explicit rather than allowing common conditions to dominate evaluation.

7.3 Practical Contribution

For fintech developers, EMFDF provides a modular architecture in which deterministic financial calculators, recommendation constraints, machine-learning personalization, and explanation are separated but interoperable. This supports implementation in digital household financial-planning applications, where household profiles, financial statements, risk assessment, education planning, insurance planning, home purchase, retirement, portfolio planning, and wealth distribution can feed a unified decision layer. For financial advisors, the framework can function as an advisory support tool rather than an autonomous replacement for professional judgment.

8. Limitations and Future Research

The study has several limitations. First, the weights and thresholds are design parameters selected for proof-of-concept evaluation; they have not yet been calibrated through a panel of certified financial planners or estimated from longitudinal household outcomes. Second, synthetic scenarios are constructed from plausible ranges but are not sampled from a national household-finance distribution. Third, the benchmark labels are scenario-defined rather than expert-adjudicated ground truth. Fourth, the machine-learning model is evaluated on data generated within the same designed environment and therefore cannot establish external generalizability. Fifth, SHAP provides feature attribution for model behavior but does not by itself establish causal explanation.

Future research should proceed in four stages. Stage 1 should use a Delphi process, AHP, or best-worst method with financial-planning experts to calibrate component weights and thresholds. Stage 2 should test the indices using anonymized real household profiles and compare EMFDF recommendations with independent professional recommendations using agreement statistics and ranking metrics such as weighted kappa, Spearman correlation, precision@k, and NDCG@k. Stage 3 should conduct a controlled user experiment comparing a conventional calculator, an AI recommendation without explanation, and EMFDF with explanation. Outcomes can include decision quality, perceived transparency, appropriate trust, recommendation acceptance, and financial self-efficacy. Stage 4 should evaluate longitudinal outcomes after real use, including changes in emergency reserves, debt burden, goal funding, and financial stress.

A future AI-agent architecture can build on the present framework by using a large language model only as an interaction and orchestration layer. Numerical calculations should be delegated to validated financial tools, recommendations should pass through the constrained decision engine, and explanations should be grounded in computed indices and model attributions. This arrangement would reduce hallucination risk and preserve traceability of financial advice.

9. CONCLUSION

This study proposes the Explainable Multi-Index Financial Decision Framework (EMFDF), an integrated household financial decision framework for personalized household financial planning. The framework integrates a Financial Health Index, Goal-Gap Index, Risk Capacity Index, and Financial Action Priority Score with explicit financial-safety constraints, a machine-learning surrogate, and SHAP-based explanations. Simulation using 1,000 scenario-stratified synthetic household profiles demonstrates strong internal performance: 92.7% agreement between the recommendation engine and the scenario benchmark, 89.0% machine-learning holdout fidelity/accuracy, macro-F1 of 0.893, macro-AUC of 0.988, and fully consistent monotonicity in the two tested directional conditions. SHAP results further indicate that model behavior is dominated by financially relevant variables such as overall goal gaps, debt burden, education gaps, emergency reserves, and retirement gaps.

The contribution of the study is not a claim that synthetic validation proves real-world advisory effectiveness. Rather, it demonstrates a transparent and testable design for integrating household financial condition, future goals, risk capacity, recommendation prioritization, and explainable AI. The framework provides a foundation for a subsequent program of expert calibration, external validation, user experimentation, and longitudinal testing. In this form, the EMFDF can be positioned as a methodological and design-science contribution to the intersection of household finance, fintech, intelligent decision support, and explainable artificial intelligence.

REFERENCES

- Bai, R. (2023). Impact of financial literacy, mental budgeting and self control on financial wellbeing: Mediating impact of investment decision making. PLOS ONE, 18(11), e0294466. https://doi.org/10.1371/journal.pone.0294466
- Carton, F., Xiong, H., & McCarthy, J. B. (2024). Human-centred factors of decision-making for financial resilience. Journal of Decision Systems, 33(sup1), 311–321. https://doi.org/10.1080/12460125.2024.2354644
- Chatterjee, S., & Fan, L. (2023). Surviving in financial advice deserts: Limited access to financial advice and retirement planning behavior. International Journal of Bank Marketing, 41(1), 70–106. https://doi.org/10.1108/IJBM-01-2022-0022
- Choung, Y., Chatterjee, S., & Pak, T. Y. (2023). Digital financial literacy and financial well-being. Finance Research Letters, 58, 104438. https://doi.org/10.1016/j.frl.2023.104438
- Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. Artificial Intelligence Review, 57, 216. https://doi.org/10.1007/s10462-024-10854-8
- Han, J., & Ko, D. (2025). Trust formation, error impact, and repair in human–AI financial advisory: A dynamic behavioral analysis. Behavioral Sciences, 15(10), 1370. https://doi.org/10.3390/bs15101370
- Kass-Hanna, J., Lyons, A. C., & Liu, F. (2022). Building financial resilience through financial and digital literacy in South Asia and Sub-Saharan Africa. Emerging Markets Review, 51, 100846. https://doi.org/10.1016/j.ememar.2021.100846
- Liu, T., Fan, M., Li, Y., & Yue, P. (2024). Financial literacy and household financial resilience. Finance Research Letters, 63, 105378. https://doi.org/10.1016/j.frl.2024.105378
- Liu, Z., & Chen, J.-K. (2024). Financial resilience in China: Conceptual framework, risk and protective factors, and empirical evidence. Journal of Family and Economic Issues, 45(4), 852–875. https://doi.org/10.1007/s10834-023-09943-7
- Mahdzan, N. S., Sabri, M. F., Husniyah, A. R., Magli, A. S., & Chowdhury, N. T. (2023). Digital financial services usage and subjective financial well-being: Evidence from low-income households in Malaysia. International Journal of Bank Marketing, 41(2), 395–427. https://doi.org/10.1108/IJBM-06-2022-0226
- Mirabile, M., Corazza, G. E., & Alonso-Moral, J. M. (2026). Trust in human–AI collaboration in finance: A bibliometric–systematic literature review. AI & Society, 41, 7625–7654. https://doi.org/10.1007/s00146-026-03049-y
- Pak, T. Y., Chatterjee, S., & Choung, Y. (2026). Navigating digital finance: How digital financial literacy influences digital financial services use and benefits in Korea. Financial Innovation, 12, 42. https://doi.org/10.1186/s40854-025-00820-w
- Peng, G., & Liu, F. (2024). Digital inclusive finance and the resilience of households involved in financial markets. Finance Research Letters, 69, 106288. https://doi.org/10.1016/j.frl.2024.106288
- Storey, V. C., Hevner, A. R., & Yoon, V. Y. (2024). The design of human-artificial intelligence systems in decision sciences: A look back and directions forward. Decision Support Systems, 182, 114230. https://doi.org/10.1016/j.dss.2024.114230
- Sunder, A., Palmer, L., Chatterjee, S., & Goetz, J. (2024). Benefits of consistent and comprehensive financial advice during the Great Recession. Journal

of Behavioral and Experimental Finance, 41, 100900. https://doi.org/10.1016/j.jbef.2024.100900

- Calcagnini, G., Favaretto, F., Giombini, G., & Tramontana, F. (2024). Household financial fragility, debt and income in a dynamic model. Computational Economics. https://doi.org/10.1007/s10614-024-10698-y
- Kim, J. H., Lee, Y., Kim, W. C., & Fabozzi, F. J. (2022). Goal-based investing based on multi-stage robust portfolio optimization. Annals of Operations Research, 313, 1141–1158. https://doi.org/10.1007/s10479-021-04473-7
- Sabharwal, R., Miah, S. J., Wamba, S. F., & Cook, P. (2025). Extending application of explainable artificial intelligence for managers in financial organizations. Annals of Operations Research, 354, 309–339. https://doi.org/10.1007/s10479-024-05825-9
- Simonato, J.-G. (2023). Maximizing the probability to reach the goal: An exploration exercise in goal-based wealth management. The Journal of Portfolio Management, 49(5), 189–207. https://doi.org/10.3905/jpm.2023.1.469
- Sinha, P., & Irala, L. R. (2025). Prepared for retirement? Think again: A systematic review and future research agenda. Management Review Quarterly, 75, 1–38. https://doi.org/10.1007/s11301-024-00437-w
- Weber, P., Carl, K. V., & Hinz, O. (2024). Applications of explainable artificial intelligence in finance: A systematic review of finance, information systems, and computer science literature. Management Review Quarterly, 74, 867–907. https://doi.org/10.1007/s11301-023-00320-0.