mathematics

MDPI

Article

A Hybrid Structural Modelling and Explainable Machine Learning Framework for Human–AI Strategic Decision Support Systems

Mostafa Aboulnour Salem

Deanship of Development and Quality Assurance, King Faisal University, P.O. Box 380, Al-Ahsa 31982, Saudi Arabia; masalem@kfu.edu.sa

Abstract

Hybrid frameworks combine structural modelling, machine learning, and explainable artificial intelligence (AI) to evaluate human–AI Strategic Decision Support Systems (SDSSs). However, their methodological reliability requires systematic assessment. This study evaluates a hybrid framework integrating partial least squares path modelling (PLS-PM), predictive machine learning, and explainable AI. It uses a reproducible Monte Carlo simulation based on the Millennial Leader–AI Agent Resistance (MLAAR) model. The model includes six socio-technical antecedents, strategic decision-making effectiveness, and institutional resilience. The experiment uses a $3 \times 2 \times 2$ factorial design. It examines three sample sizes (N = 200, 500, and 1000), two antecedent-correlation levels ($\rho = 0.0$ and 0.3), and two functional forms (linear and nonlinear). Each condition includes 100 replications, producing 1200 synthetic datasets. Structural recovery, bootstrap inference, predictive accuracy, and feature importance are evaluated against known population parameters. Mean relative coefficient bias increased in magnitude from −11.65% under independent antecedents to −29.97% under correlated antecedents. Correct identification of the strongest population antecedent increased from 70.0% to 96.3% as sample size increased. However, structural and SHAP predictor rankings agreed exactly only 22.3% of the time. Indirect-effect confidence-interval coverage declined from 65.8% to 9.8%, despite consistent statistical detection. Ordinary least squares regression achieved the highest average cross-validated predictive accuracy among the evaluated algorithms. These findings show that predictive accuracy, statistical detection, and predictor-ranking agreement do not guarantee accurate structural recovery. The study provides a reproducible framework for evaluating hybrid explanatory–predictive methods. Independent empirical validation remains necessary before applying the framework to organisational decision-making.

Keywords: agentic artificial intelligence; Strategic Decision Support Systems; managerial resistance; structural equation modelling; explainable machine learning; decision authority; human–AI collaboration; institutional resilience; Industry 5.0

MSC: 65C05; 62H25; 62F40; 62J05; 68T01; 68T05; 90B50

Check for updates

Academic Editors: Xiaobing Yu and Mei Cai

Received: 20 September 2026
Revised: 21 October 2026
Accepted: 22 October 2026
Published: 24 October 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Artificial intelligence (AI) is transforming Strategic Decision Support Systems (SDSSs). Intelligent agents can analyse complex information, evaluate alternatives, generate recommendations, and perform decision-related tasks with increasing autonomy [1–4]. These

capabilities can improve analytical efficiency and organisational responsiveness. However, they also raise concerns about managerial authority, transparency, trust, accountability, and human oversight [5,6]. These challenges are particularly relevant to Industry 5.0. This paradigm emphasises human-centricity, organisational resilience, and the responsible integration of intelligent technologies [2,7].

Consequently, human–AI strategic decision-making requires analytical frameworks. These provide clear theoretical explanations and reliable predictive evidence. The process involves interconnected technological, behavioural, and organisational factors. AI Agent Autonomy (AIAA) refers to intelligent agents' operational independence. In contrast, AI Decision Opacity (AIDO) refers to the difficulty of understanding algorithmic recommendations [8,9].

Perceived Decision-Authority Threat (PDAT) reflects concerns about the potential loss of managerial discretion. In contrast, Human–AI Trust (HAIT), AI Governance and Human Oversight (AIGO), and Human-Centric AI Orientation (HCAIO) reflect conditions related to confidence, accountability, and meaningful human participation [10–13]. These six antecedents form the theoretical foundation of Millennial Leader–AI Agent Resistance (MLAAR). They also provide a basis for examining its proposed relationships with Strategic Decision-Making System Effectiveness (SDMSE) and Institutional Resilience (IR).

Recent research has increasingly combined structural modelling, machine learning, and explainable AI to examine complex socio-technical relationships. Partial least squares path modelling (PLS-PM) estimates theoretically specified structural associations. Machine-learning algorithms evaluate predictive performance and can capture nonlinear relationships [14–18].

SHapley Additive exPlanations (SHAP) help explain how predictors contribute to model predictions [19,20]. These methods provide complementary analytical perspectives. However, they measure different statistical quantities. Structural coefficients estimate model-dependent associations. Predictive metrics assess generalisation accuracy. SHAP values explain predictions generated by fitted models. Agreement between these results does not necessarily indicate accurate recovery of the underlying population relationships.

Despite advances in hybrid explanatory–predictive frameworks, four methodological gaps remain. First, measurement errors, ordinal indicators, and correlated antecedents may affect structural coefficients. These factors can introduce systematic estimation bias [21–23]. Second, superior predictive performance does not confirm the underlying functional form. Linear models may outperform nonlinear alternatives under certain sample sizes and modelling conditions [24,25]. Third, SHAP importance and structural coefficients measure different aspects of predictor relationships.

Therefore, agreement between their rankings can be difficult to interpret. This issue is particularly important when predictors are correlated [26,27]. Fourth, statistically significant direct and indirect associations may coexist with biased estimates. Confidence intervals may also inadequately cover the true population parameters [23,28]. These limitations highlight the need for systematic methodological evaluation. Such evaluation is essential before drawing substantive conclusions from integrated explanatory and predictive analyses.

Monte Carlo simulation provides an appropriate approach to examining these methodological concerns. This is possible because the generating relationships and population parameters are known [26,27], unlike observational studies, where the true underlying coefficients are generally unknown [28]. A controlled simulation permits direct assessment of estimation bias and root mean squared error (RMSE). It also assesses directional recovery, predictor-ranking accuracy, bootstrap confidence-interval coverage, and out-of-sample predictive performance.

Accordingly, the present study evaluates the reliability of a hybrid PLS-PM–machine learning–SHAP framework. The MLAAR model is used as a theoretically specified simulation population rather than an empirically validated representation of managerial behaviour. The experiment employs a $3 \times 2 \times 2$ factorial design comprising three sample sizes (N = 200, 500, and 1000), two antecedent-correlation conditions (ρ = 0.0 and 0.3), and two generating functional forms (linear and nonlinear). Each of the 12 conditions includes 100 independent replications, producing 1200 synthetic datasets.

The analytical pipeline integrates reflective Mode-A PLS-PM, bootstrap estimation of structural and indirect associations, five-fold cross-validated prediction using ordinary least squares regression (OLS), gradient boosting, and artificial neural networks (ANNs), and SHAP attribution based on held-out observations. The study addresses five research questions:

RQ1. How accurately does PLS-PM recover the MLAAR population coefficients across different sample sizes, antecedent correlations, and functional forms?
RQ2. How accurately do PLS-PM and SHAP recover the true population predictor rankings?
RQ3. How accurately do the methods recover the direct and indirect associations between MLAAR and IR?
RQ4. How does the cross-validated predictive performance of OLS, gradient boosting, and ANN differ under linear and nonlinear conditions?
RQ5. What methodological limitations arise when integrating PLS-PM, machine learning, and SHAP in human–AI SDSSs?

The study makes three main contributions. First, it develops a reproducible simulation framework with known population parameters. This framework evaluates structural coefficient recovery and mediation inference. Second, it distinguishes population-ranking recovery from agreement between PLS-PM and SHAP importance. This distinction clarifies the limitations of convergence between explanatory and predictive methods.

Third, the study systematically compares linear and nonlinear predictive methods under controlled simulation conditions. It examines why predictive accuracy cannot independently confirm structural validity. Together, these contributions provide methodological guidance for evaluating hybrid analytical frameworks.

The remainder of the article is organised as follows. Section 2 presents the theoretical background and hypotheses. Section 3 describes the population-generating model, Monte Carlo simulation design, and analytical procedures. Section 4 reports the structural, predictive, mediation, and explainability results. Section 5 discusses the main findings. Section 6 presents the study's implications. Section 7 discusses limitations and future research directions. Finally, Section 8 presents the conclusions.

2. Literature Review and Hypothesis Development

2.1. Technological Conditions and Resistance to Agentic AI

Agentic artificial intelligence (AI) is transforming Strategic Decision Support Systems (SDSSs). Intelligent systems can analyse alternatives, generate recommendations, and perform delegated decision-related tasks. These capabilities may improve efficiency and responsiveness. However, they also raise concerns about managerial control, transparency, and accountability [3,13].

User-resistance theory can explain managerial resistance to AI. In this theory, resistance is a behavioural response to a perceived threat created by a new system, rather than a simple lack of acceptance [29,30]. Status quo bias theory adds that users weigh the expected gains of change against switching costs and possible losses, and that perceived losses usually weigh more heavily than equal gains [29]. Research on algorithm aversion supports this view.

People often prefer human judgement to algorithmic advice, particularly after observing algorithmic errors and in tasks that seem subjective or consequential [6,31–33]. Agentic AI intensifies these concerns. Unlike conventional decision aids, AI agents can plan, act, and coordinate tasks with limited human input [10,34]. Human and machine agency therefore become intertwined in organisational decisions [35]. This creates new uncertainty about who controls a decision and who is accountable for its outcome [36,37]. The MLAAR framework applies this reasoning to Millennial leaders. This cohort entered leadership roles alongside digital technologies. However, generational research shows that work values differ across cohorts [38–40]. Familiarity with technology therefore does not imply unconditional acceptance of delegated authority for AI decisions.

Within the Millennial Leader–AI Agent Resistance (MLAAR) framework, two technological factors are particularly relevant. These are AI Agent Autonomy (AIAA) and AI Decision Opacity (AIDO). AIAA refers to the extent to which AI agents perform tasks without continuous human supervision. Greater autonomy may limit opportunities for managerial intervention. It may also increase uncertainty about responsibility for decisions [28–32].

AIDO refers to the difficulty of understanding how AI agents process information and generate recommendations. Limited transparency may reduce decision-makers' confidence in evaluating AI-generated outputs. This concern matters most when decisions involve uncertainty and institutional accountability [33–36].

AIAA is expected to increase resistance because people perceive a loss of control. Models of automation levels show that higher autonomy shifts information acquisition, decision selection, and action implementation from humans to machines [41–43]. Each shift reduces opportunities for human intervention. Research on decision automation indicates that people respond more negatively to automated decisions than to augmented decisions in which humans keep final authority [37].

Studies of algorithm aversion also show that people accept imperfect algorithms more readily when they can modify the output [44]. Removing this possibility should therefore increase resistance. In status quo-bias terms, leaders weigh lost control as a switching cost [29]. AIDO is expected to increase resistance through uncertainty and accountability concerns. Opacity arises from technical complexity, commercial secrecy, and the mismatch between high-dimensional machine learning and human reasoning [45].

When leaders cannot understand a recommendation, they cannot easily verify it or justify it to stakeholders [46,47]. Experimental evidence shows that explainability increases trust in and acceptance of AI [48,49]. Algorithmic transparency also improves the perceived trustworthiness of automated decisions [50]. Therefore, opacity increases resistance, mainly because leaders remain accountable for decisions they cannot fully explain.

Although related, autonomy and opacity represent different technological characteristics. Autonomy concerns who acts. Opacity concerns whether people can understand the reasoning behind an action [51]. A highly autonomous system can be transparent, and an advisory system can be opaque. Both are expected to increase resistance, but through separate mechanisms. Based on these theoretical considerations, the following hypotheses are proposed:

H1. *AIAA is positively associated with MLAAR.*

H2. *AIDO is positively associated with MLAAR.*

2.2. Human–AI Interaction, Governance, and Human-Centricity

Managerial resistance to agentic AI stems from both technological and organisational factors. Perceptions of decision authority, trust, and institutional safeguards also play

important roles. Moreover, the MLAAR framework includes four additional antecedents: Perceived Decision-Authority Threat (PDAT), Human–AI Trust (HAIT), AI Governance and Human Oversight (AIGO), and Human-Centric AI Orientation (HCAIO). PDAT is a risk factor, whereas HAIT, AIGO, and HCAIO are mitigating factors.

PDAT reflects concerns that AI agents may reduce managerial discretion, professional influence, or control over strategic decisions [1,32,52]. It is expected to be the most proximal antecedent of resistance. User-resistance theory holds that resistance arises when people perceive a threat to their position, status, or power [30]. Lapointe and Rivard showed that perceived threats to individuals' and groups' power trigger resistance at several organisational levels [30,31].

Research on algorithmic management also shows that algorithms can reshape workplace control and provoke resistance from those whose authority is affected [53]. Managers may see delegated AI decision-making as a threat to their professional role [36–39]. AIAA and AIDO describe technology. In contrast, PDAT describes how a leader appraises the personal consequences of the technology. Such threat appraisals are expected to predict resistance more directly than system characteristics [29,30].

HAIT reflects confidence in the reliability and appropriate functioning of AI-supported recommendations. Trust is the willingness to be vulnerable to another party whose actions cannot be fully monitored [54]. It therefore matters most when AI agents act autonomously. Trust in automation guides reliance when users cannot fully understand a system [42]. Reviews of empirical research identify trust as a central determinant of AI adoption in organisations [8].

Recent studies also show that higher trust increases the weight that decision-makers give to AI advice [55,56]. Calibrated trust can reduce unnecessary resistance while preserving critical human judgement [42]. HAIT is therefore expected to be negatively associated with MLAAR. AIGO refers to institutional arrangements for monitoring AI-supported decisions, assigning responsibility, and allowing human intervention and override [57–59]. Responsible AI governance translates ethical principles into structural, procedural, and relational practices within organisations [60].

Such safeguards reduce uncertainty about accountability when an AI agent errs [61]. Monitoring and overriding algorithmic outputs also reduce algorithm aversion [44]. However, governance is expected to act on resistance partly through trust and perceived control. Its unique association with MLAAR is therefore expected to be negative but weaker than that of HAIT. HCAIO reflects an organisational commitment to supporting human expertise rather than replacing managerial judgement. It aligns with Industry 5.0, which places human well-being, agency, and resilience at the centre of technological integration [2,7,62].

Human-centred AI aims to combine high levels of automation with high levels of human control [3]. Augmentation-oriented uses of AI also tend to pose less of a threat than automation-oriented uses [36,37]. When leaders see that AI is designed to augment their judgement, delegation becomes less threatening. HCAIO is therefore expected to be negatively associated with MLAAR. Together, these four constructs represent complementary but theoretically distinct dimensions of human–AI decision support. Based on these theoretical considerations, the following hypotheses are proposed:

H3. *PDAT is positively associated with MLAAR.*

H4. *Human–AI Trust (HAIT) is negatively associated with MLAAR.*

H5. *AI Governance and Human Oversight (AIGO) is negatively associated with MLAAR.*

H6. *Human-Centric AI Orientation (HCAIO) is negatively associated with MLAAR.*

2.3. MLAAR, SDMSE, and IR

The effects of managerial resistance may extend beyond individual attitudes towards AI. Strategic Decision-Making System Effectiveness (SDMSE) refers to a human–AI system's ability to support informed, timely, and contextually appropriate decisions [9,13,63]. Strategic decision effectiveness depends on the quality of information processing and on how fully available analytical resources are used [64].

Resistance can reduce SDMSE in three ways. Leaders may ignore or override accurate AI recommendations [5,37]. They may restrict the information and tasks they delegate to AI agents. They may also delay decisions while seeking human confirmation. Hybrid intelligence is effective only when human and machine contributions are combined [9,11]. Resistance weakens this combination. Greater MLAAR is therefore expected to be associated with lower SDMSE.

Institutional Resilience (IR) refers to an organisation's ability to anticipate challenges, cope with disruptions, and adapt to change [65,66]. Duchek describes resilience as a set of capabilities for anticipating, coping with, and adapting to change [56]. Each capability depends on effective decision-making. Anticipation requires accurately interpreting early signals. Coping requires timely decisions under pressure. Adaptation requires learning from decision outcomes.

Effective human–AI decision systems can strengthen these capabilities by improving information processing, coordination, and responsiveness. SDMSE is therefore expected to be positively associated with IR. Combining these arguments implies an indirect relationship between MLAAR and IR through SDMSE. Resistance is expected to weaken resilience partly because it reduces decision effectiveness.

However, resistance may also affect resilience through other channels, such as slower organisational learning and slower technological adaptation. The model therefore does not assume complete mediation. It retains a direct pathway between MLAAR and IR, in line with current guidance on testing indirect effects alongside direct effects [67–70]. Based on these theoretical considerations, the following hypotheses are proposed:

H7. *MLAAR is negatively associated with SDMSE.*

H8. *SDMSE is positively associated with IR.*

H9. *MLAAR has a negative indirect association with IR through SDMSE.*

The nine hypotheses provide the theoretical foundation for the Monte Carlo simulation. Each hypothesis is represented by a prespecified population coefficient (Tables 1 and 2). These coefficients are nominal design values, not empirical estimates. Their signs follow the hypothesised directions, and their relative sizes follow the arguments above.

Table 1. Prespecified Population Coefficients for the MLAAR Antecedents.

Hypothesis	Structural Relationship	γ
H1	AIAA → MLAAR	0.240
H2	AIDO → MLAAR	0.260
H3	PDAT → MLAAR	0.370
H4	HAIT → MLAAR	−0.180
H5	AIGO → MLAAR	−0.090
H6	HCAIO → MLAAR	−0.210

Table 2. Prespecified Downstream Structural and Mediation Parameters.

Hypothesis	Structural Relationship	Parameter	Population Value
H7	MLAAR → SDMSE	a	−0.400
H8	SDMSE → IR	b	0.450
H9	MLAAR → SDMSE → IR	ab	−0.180
Additional direct pathway	MLAAR → IR	c'	−0.100
Total association	MLAAR → IR	c' + ab	−0.280

PDAT receives the largest coefficient (0.37) because it is the most proximal, person-level appraisal of threat. AIDO (0.26) and AIAA (0.24) receive smaller positive coefficients because they describe system characteristics whose influence is expected to operate partly through threat appraisals. Among the mitigating factors, HCAIO (−0.21) and HAIT (−0.18) receive similarly sized, moderate negative coefficients. AIGO (−0.09) has the smallest coefficient because it is expected to operate largely through trust and perceived control.

All six values lie in the small-to-medium range of standardised effects [68]. This range avoids both trivial and implausibly large effects, providing a realistic test of estimator accuracy. The downstream coefficients are MLAAR → SDMSE = −0.40, SDMSE → IR = 0.45, and MLAAR → IR = −0.10. The resulting indirect association is −0.18. It accounts for about 64% of the total association (−0.28), which represents partial mediation. Table A4 (Appendix C) summarises the theoretical basis, mechanism, key sources, and population value of each hypothesis.

Figure 1 illustrates the theory-derived MLAAR framework and its prespecified population coefficients. The subsequent Monte Carlo analysis evaluates the accuracy of the hybrid analytical framework. It examines how reliably structural modelling and explainable machine-learning methods recover these known population relationships.

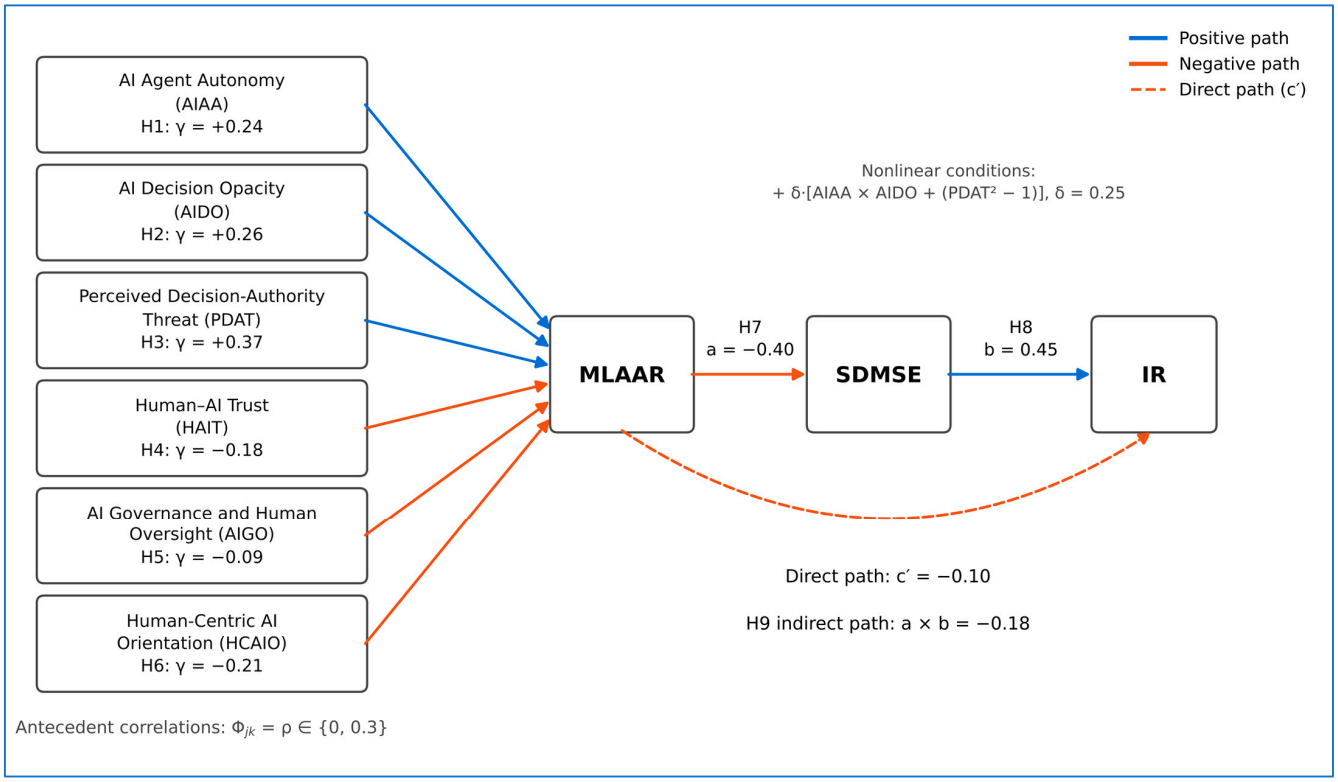


Figure 1. The theory-derived MLAAR framework.

2.4. Research Gap Synthesis and Conceptual Contribution

Existing research on agentic artificial intelligence (AI), algorithmic decision-making, and human–AI collaboration has improved our understanding of autonomy, transparency, managerial authority, trust, and organisational governance [41–45]. However, four interconnected research gaps remain. These gaps are particularly relevant to developing and evaluating hybrid analytical frameworks for Strategic Decision Support Systems (SDSSs).

First, the socio-technical integration gap concerns the limited examination of technological, human, and organisational factors within a unified structural framework. Previous studies have examined these factors separately or in different combinations. However, estimating their relationship can be challenging. Measurement error and correlations between antecedents may affect estimation accuracy. The Millennial Leader–AI Agent Resistance (MLAAR) framework addresses this conceptual gap. It integrates six theoretically distinct antecedents into a common structural model.

Second, the organisational outcome gap concerns the relationships between managerial resistance, Strategic Decision-Making System Effectiveness (SDMSE), and Institutional Resilience (IR). These constructs are important for understanding human–AI strategic decision-making. However, estimating their direct and indirect associations requires careful model specification. It also requires reliable statistical inference. Therefore, the proposed framework includes a partial-mediation structure. It does not assume that SDMSE fully mediates the relationship between MLAAR and IR [23,67].

Third, the methodological reliability gap concerns integrating structural modelling, predictive machine learning, and explainable AI. Partial least squares path modelling (PLS-PM), predictive algorithms, and SHapley Additive exPlanations (SHAP) provide complementary analytical evidence [22]. However, their outputs measure different statistical quantities. Structural coefficient accuracy, predictive performance, and feature-importance agreement do not measure the same aspect of model validity. Moreover, statistically significant associations may coexist with biased estimates. Confidence intervals may also fail to cover the true population parameters adequately [23,28].

Fourth, the validation and interpretation gap concerns evaluating hybrid analytical frameworks when the true population relationships are unknown. Agreement between structural coefficients and SHAP rankings does not confirm accurate recovery of the true predictor ordering. Similarly, strong predictive performance does not establish structural accuracy. It also does not provide sufficient evidence to justify specific managerial interventions. Known-parameter simulation offers a controlled approach to examining these limitations. It allows researchers to assess analytical reliability before drawing substantive conclusions.

The present study addresses these four gaps through a reproducible Monte Carlo simulation. It uses the theory-derived MLAAR framework as a prespecified population model. The main objective is to evaluate the reliability of structural coefficient recovery, partial-mediation inference, predictive performance, and SHAP-based predictor rankings. We conduct these evaluations under controlled generating conditions. The study also distinguishes the recovery of known population relationships from agreement between estimated analytical outputs. This distinction provides a clearer basis for interpreting convergence between explanatory and predictive methods [27,71].

2.5. Related Methodological Work on Hybrid Structural–Predictive Frameworks

Hybrid frameworks that combine structural models with machine learning are now common in information systems and management research. Early applications used a two-stage SEM–ANN approach. Structural equation modelling first identified significant predictors. An artificial neural network then ranked these predictors and captured nonlin-

ear relationships [72,73]. Leong et al. applied this approach to predict resistance to mobile wallets, a context close to the present study [74].

Later work added explainable AI to these pipelines. Researchers now use SHAP values to rank predictors and compare predictive importance with structural coefficients [19,20]. Related decision-support studies have also combined structural modelling, educational data mining, natural language processing, and explainable machine learning [12–15]. A parallel literature asks how such outputs should be interpreted. Shmueli distinguished explanatory modelling, which tests causal theory, from predictive modelling, which maximises out-of-sample accuracy [17].

Hofman et al. argued that the two goals should be integrated but not confused [74]. Guidelines for PLS-SEM now recommend separate assessment of explanatory quality and out-of-sample predictive power [16,75]. Comparative studies also show that flexible machine-learning models do not always outperform regression models on structured data [24,25]. Simulation research has examined the accuracy of PLS-PM itself. Reinartz et al. and Goodhue et al. compared PLS with covariance-based SEM (CB-SEM) using known population models [76,77]. Both studies reported that PLS tends to underestimate structural paths when indicators contain measurement error.

Sarstedt et al. showed that the direction and size of the bias depend on whether the population model is composite-based or factor-based [78]. Rönkkö et al. questioned using PLS to infer latent variables [79]. Consistent PLS was developed to correct this attenuation [48,49]. Yuan et al. showed that the choice of estimator also affects the precision of mediation estimates [53]. Research on explainable AI raises related concerns. Shapley values distribute a model's prediction among its inputs [19,20]. However, the results depend on how absent predictors are represented, especially when predictors are correlated [26,27]. Kumar et al. showed that Shapley-based importance does not necessarily reflect the structure of the data-generating process [80]. Dependence-aware alternatives have therefore been proposed [26,27].

Table A3 (Appendix C) summarises representative studies from these strands. Few studies have evaluated all three components together. Most hybrid applications analysed a single empirical dataset with unknown population parameters. They could therefore assess agreement between methods, but not accuracy against a known truth. In contrast, simulation studies of PLS-PM did not examine predictive models or SHAP rankings. The present study combines these strands. It evaluates structural recovery, mediation inference, predictive accuracy, and SHAP rankings against the same known population model.

3. Materials and Methods

3.1. Research Design, Population, and Data Generation

This study employed a factorial Monte Carlo simulation to evaluate the methodological reliability of a hybrid structural-modelling and explainable machine-learning framework for human–AI Strategic Decision Support Systems (SDSSs). The Millennial Leader–AI Agent Resistance (MLAAR) framework and its nine theory-derived hypotheses (H1–H9) served as the conceptual foundation of the data-generating model. The study generated synthetic observations from a fully specified population model and assessed whether the analytical pipeline recovered its known relationships.

The population comprised six exogenous constructs: AI Agent Autonomy (AIAA), AI Decision Opacity (AIDO), Perceived Decision-Authority Threat (PDAT), Human–AI Trust (HAIT), AI Governance and Human Oversight (AIGO), and Human-Centric AI Orientation (HCAIO). The antecedents were generated jointly from a multivariate normal distribution:

$$\xi \sim \mathcal{N}_6(0, \Phi) \quad (1)$$

where ξ denotes the six antecedents and Φ is an equicorrelation matrix with diagonal elements equal to 1 and off-diagonal elements equal to ρ.

MLAAR was generated using

$$\text{MLAAR}^* = \gamma^T \xi + \delta\, NL(\xi) + \zeta_1 \tag{2}$$

where γ contains the prespecified antecedent coefficients, δ controls the nonlinear contribution, and ζ represents normally distributed residual variation.

To examine functional-form sensitivity, the nonlinear component was defined as:

$$NL(\xi) = \frac{\xi_{\text{AIAA}}\xi_{\text{AIDO}} + \left(\xi^2_{\text{PDAT}} - 1\right)}{\sqrt{3 + 5\rho^2}} \tag{3}$$

The nonlinear parameter was set to δ = 0 for linear conditions and δ = 0.25 for nonlinear conditions. The residual variance was specified as $max(1 - \gamma^{\text{T}}\Phi\gamma - \delta^2, 0.05)$. MLAAR was subsequently standardised within each generated dataset. Consequently, the assigned coefficients represent nominal generating parameters, and comparisons with estimated standardised coefficients must be interpreted considering this standardisation and the measurement process [53,80].

The downstream population model incorporated Strategic Decision-Making System Effectiveness (SDMSE) and Institutional Resilience (IR):

$$\text{SDMSE} = a\,\text{MLAAR} + \zeta_2 \tag{4}$$

$$\text{IR} = b\,\text{SDMSE} + c'\,\text{MLAAR} + \zeta_3 \tag{5}$$

The SDMSE residual variance was set to $(1 - a^2)$, while the IR residual variance was set to $(1 - (b^2 + c'^2 + 2abc'))$. The model retained the direct MLAAR–IR pathway to avoid imposing complete mediation.

The Monte Carlo experiment followed *a* 3 × 2 × 2 factorial design with three sample sizes N = 200, 500, and 1000, two antecedent-correlation conditions ρ = 0.0 and 0.3 and two generating functional forms (linear and nonlinear). Each of the 12 conditions comprised 100 independently seeded replications, yielding 1200 synthetic datasets. (see Table 3)

To ensure reproducibility, each experimental condition and replication was assigned a deterministic random seed:

$$s = 20{,}261{,}014 + 100{,}000 i_N + 10{,}000 i_\rho + 1000 i_f + r \tag{6}$$

where i_N, i_ρ, and i_f denote zero-based factor-level indices, and r denotes the replication index.

Equations (1)–(8) were developed for this study. They define the data-generating process. Equation (1) specifies the antecedent distribution. Equations (2)–(5) describe the structural relationships. Equation (6) defines the random seed assignment. Equations (7) and (8) specify the measurement and ordinal-discretisation procedures. These equations follow general simulation principles described by Morris et al. [46]. They also draw on the reflective measurement approach used in PLS path modelling [48]. However, the parameter values and functional forms were specifically developed for the MLAAR population model.

Table 3. Monte Carlo Experimental Design and Computational Specifications.

Component	Specification
Research design	Factorial Monte Carlo simulation
Population model	Theory-derived MLAAR framework
Sample sizes	200, 500, 1000
Antecedent correlation	ρ = 0.0 and 0.3
Functional forms	Linear and nonlinear
Experimental conditions	12
Replications per condition	100
Total simulated datasets	1200
Latent constructs	9
Indicators per construct	4
Response categories	Five-point ordinal
Structural estimator	Reflective Mode-A PLS-PM
Bootstrap resamples	100 per replication
Predictive validation	Shuffled five-fold cross-validation
Predictive algorithms	OLS, gradient boosting, ANN
Explainability	Exact interventional Shapley attribution
Base random seed	20,261,014

3.2. Measurement Model and Construct Operationalisation

The measurement model comprised nine reflective constructs: the six MLAAR antecedents, MLAAR, SDMSE, and IR. Each construct was represented by four indicators, producing 36 observed variables per simulated dataset. This design enabled evaluation of the analytical pipeline using imperfect ordinal measurements rather than directly observed latent population values.

For indicator (r) of construct (j), the underlying continuous response was generated as:

$$y_{jr}^{*} = \lambda_r \eta_j + \sqrt{1 - \lambda_r^2}\, \varepsilon_{jr} \tag{7}$$

where η_j denotes the generating latent construct, λ_r represents its indicator loading, and ε_{jr} is an independent standard-normal measurement disturbance. The fixed loading pattern was (0.72, 0.78, 0.83, 0.88).

Continuous responses were converted into five ordinal categories using thresholds of −1.90, −0.90, 0.35, and 1.45:

$$y_{jr} = 1 + \sum_{k=1}^{4} \mathbb{I}\left(y_{jr}^{*} > \tau_k\right) \tag{8}$$

where $\mathbb{I}$ denotes the indicator function and τ_k represents the corresponding threshold.

Measurement quality was examined through standardised indicator loadings, Cronbach's alpha, composite reliability (CR), and average variance extracted (AVE). For standardised reflective indicators, AVE was calculated as:

$$\mathrm{AVE}_j = \frac{1}{m_j} \sum_{r=1}^{m_j} \lambda_{jr}^2 \tag{9}$$

where m_j denotes the number of indicators per construct. Composite reliability was calculated using the corresponding standardised loadings and error variances. Reliability values of 0.70 and AVE values of 0.50 were used as conventional screening criteria [16,81,82].

Discriminant validity was assessed using the heterotrait–monotrait ratio (HTMT). A threshold of 0.85 was used for screening. Potential multicollinearity between the six antecedents was examined using variance inflation factors (VIFs). Values below 3.3 were considered acceptable based on a conservative screening criterion [16,83]. Particular attention was given to AIAA, AIDO, and PDAT. These constructs represent conceptually related but distinct dimensions of AI autonomy, decision transparency, and managerial authority.

This study distinguishes two kinds of loadings. The population loadings λ_r in Equation (7) are fixed design values, identical (0.72, 0.78, 0.83, 0.88) for all nine constructs. The outer loadings reported in Section 4.1 are estimates: correlations between each ordinal indicator and the construct score produced by the fitted PLS-PM. In Mode A, the outer weights of an exogenous construct are proportional to the covariances between its indicators and the score of its adjacent endogenous construct [23,48]. For antecedent j, the population correlation with MLAAR is

$$\mathrm{Corr}\left(\xi_j, \mathrm{MLAAR}\right) = \gamma_j + \rho \sum_{k \neq j} \gamma_k \tag{10}$$

Hence, a construct with a moderate coefficient can have a near-zero marginal correlation when $\rho > 0$, and its outer weights are then determined largely by sampling error. For example, with $\rho = 0.3$, the population correlation of HCAIO with MLAAR is $-0.21 + 0.3 \times 0.60 = -0.030$.

To separate such estimation effects from the measurement model, three additional diagnostics were computed: (i) the correlation of each ordinal indicator with its own true latent variable and its largest correlation with any other latent variable, which verifies indicator coding and mapping; (ii) unit-weight (mean-of-items) composite loadings, which do not depend on the structural model; and (iii) the correlation between each PLS-PM construct score and its true latent variable. These diagnostics were computed for the illustrative dataset and for all 600 linear-condition replications.

3.3. Structural Model Estimation and Bootstrap Inference

Structural relationships were estimated using reflective Mode-A partial least squares path modelling (PLS-PM) with centroid inner weighting. The algorithm iteratively estimated measurement-block weights and construct scores, using a consistent positive orientation for the positively keyed indicators.

Following convergence, standardised construct scores were used to estimate the structural pathways through ordinary least squares regression. The estimated MLAAR model was:

$$\widehat{\mathrm{MLAAR}} = \sum_{j=1}^{6} \hat{\gamma}_j \hat{\xi}_j + \hat{\varepsilon}_1 \tag{11}$$

The downstream structural equations were estimated using the partial-mediation specification in Equations (4) and (5). The recovered indirect and total associations were calculated as:

$$\widehat{IE} = \hat{a}\hat{b}, \quad \widehat{TE} = \hat{c}' + \hat{a}\hat{b} \tag{12}$$

Estimated coefficients were compared with the corresponding nominal generating values in Tables 1 and 2. Structural recovery was evaluated through coefficient bias, relative bias, root mean squared error (RMSE), and directional accuracy. Predictor-strength

comparisons used absolute standardised coefficients, without assuming that numerical differences were statistically significant.

Direct and indirect association uncertainty was assessed using nonparametric bootstrap resampling. Within each Monte Carlo replication, 100 bootstrap samples were generated by sampling observations with replacement. The complete PLS-PM procedure was re-estimated for each bootstrap sample, and percentile-based 95% confidence intervals were constructed:

$$CI_{95\%}(\theta) = \left[Q_{0.025}\left(\hat{\theta}^*\right),\ Q_{0.975}\left(\hat{\theta}^*\right)\right] \tag{13}$$

Here, $\hat{\theta}^*$ represents a bootstrap estimate and Q denotes the corresponding empirical quantile.

Bootstrap performance was evaluated using detection probability (the proportion of intervals excluding zero) and coverage probability (the proportion containing the known generating parameter). These measures were distinguished because statistical detection does not necessarily imply accurate estimation or properly calibrated uncertainty [46,47,83].

Detection and coverage rates were estimated at 100 Monte Carlo replications per condition. Their Monte Carlo standard errors (MCSEs) were calculated using [46]:

$$\mathrm{MCSE}(\widehat{p}) = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{R}} \tag{14}$$

which is at most 5.0 percentage points for R = 100 and at most 2.5 percentage points when the four conditions at each sample size are pooled (R = 400). Because confidence-interval coverage is a primary evaluation criterion and each replication used B = 100 resamples, a bootstrap sensitivity analysis was conducted for the six linear conditions (N = 200, 500, and 1000; ρ = 0.0 and 0.3).

The same 600 datasets as in the main study (identical seeds) were re-analysed with B = 1000 percentile-bootstrap resamples, re-estimating the complete PLS-PM model in each resample. Within each replication, intervals based on the first 100 resamples were compared with intervals based on all 1000, so both bootstrap budgets were evaluated on identical datasets.

The separate illustrative dataset was analysed using 5000 bootstrap resamples. It also included paired coefficient contrasts between PDAT, AIDO, and AIAA. These illustrative contrasts were based on the fitted structural coefficients and their bootstrap distributions; their individual-dataset results were not pooled with the main Monte Carlo condition summaries.

3.4. Predictive Modelling and Model Comparison

The predictive component evaluated the ability of six socio-technical antecedents to predict MLAAR on held-out synthetic observations. To avoid leakage from full-dataset PLS-PM estimation, predictive variables were represented by prespecified arithmetic means of their four ordinal indicators:

$$x_{ij}^{(c)} = \frac{1}{4}\sum_{k=1}^{4} y_{ijk} \tag{15}$$

where $x_{ij}^{(c)}$ represents the composite score of construct j for observation i.

Three algorithms were evaluated: ordinary least squares (OLS) regression, gradient-boosted trees, and an artificial neural network (ANN), implemented in scikit-learn [84]. OLS provided an additive benchmark, while gradient boosting and the ANN permitted more flexible nonlinear prediction. (see Table 4)

Table 4. Predictive Algorithms and Model Configurations.

Algorithm	Implemented Configuration	Purpose
OLS	Linear regression with intercept	Additive benchmark
Gradient boosting	Histogram-based estimator; 300 iterations; depth 2; learning rate 0.05; L2 regularisation 1.0	Nonlinear prediction
ANN	One hidden layer with 8 neurons; L-BFGS optimiser; L2 penalty 1.0; maximum 500 iterations	Flexible nonlinear prediction

Each dataset was evaluated using shuffled five-fold cross-validation, with identical training and testing partitions across algorithms. Data-dependent standardisation was fitted using only training-fold observations and then applied to the corresponding held-out observations. The predictions from the five test folds were combined into a single out-of-fold prediction vector.

Predictive performance was evaluated through the cross-validated coefficient of determination:

$$R^2_{\mathrm{CV}} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{16}$$

where y_i and $\hat{y}_i$ represent the synthetic observed and out-of-fold predicted MLAAR composites, respectively.

Equation (16) is the standard out-of-sample coefficient of determination computed from cross-validated predictions [25,52]. For each experimental condition, cross-validated R^2 was averaged across the 100 replications. The proportion of replications in which OLS outperformed gradient boosting was also recorded.

Differences were interpreted as descriptive comparisons of the implemented algorithms rather than evidence of universal algorithmic superiority. The illustrative dataset was evaluated separately using five repetitions of five-fold cross-validation, producing 25 held-out evaluations. Its additional predictive diagnostics included RMSE and mean absolute error (MAE).

3.5. Explainable AI and Explanatory–Predictive Convergence

Explainable AI was incorporated using SHapley Additive exPlanations (SHAP) to assess the contributions of the six antecedents to gradient-boosting predictions [19]. The analysis used exact interventional Shapley attribution, evaluating all 64 possible coalitions of the six predictors.

For predictor (j), the Shapley contribution was calculated as:

$$\phi_j(x) = \sum_{S \subseteq F,\, j \notin S} w(S)[v_x(S \cup \{j\}) - v_x(S)] \tag{17}$$

$$w(S) = \frac{|S|!(p - |S| - 1)!}{p!} \tag{18}$$

where F denotes the predictor set, $p = 6$, and $v_x(S)$ represents the coalition-specific reference prediction.

SHAP explanations were generated using held-out observations from the five-fold cross-validation procedure. Each reference sample included up to 32 training observations. Up to 200 held-out observations were explained in each replication. All preprocessing

transformations were fitted using only the corresponding training fold. Global feature importance was calculated as follows:

$$I_j^{\text{SHAP}} = \frac{1}{M}\sum_{i=1}^{M}|\phi_{ij}| \tag{19}$$

where M denotes the number of explained observations. Mean absolute SHAP importance was used to rank the six antecedents.

Three ranking comparisons were conducted. First, estimated PLS-PM coefficient rankings were compared with the known population ranking. Second, SHAP importance rankings were compared with the population ranking. Third, estimated PLS-PM rankings were compared with SHAP rankings. Spearman's rank correlation was used to measure agreement between the rankings. The analysis also recorded how often PDAT was correctly identified as the strongest antecedent. In addition, it measured the frequency of complete ranking matches.

These comparisons distinguish population-ranking recovery from agreement between structural and predictive methods. PLS-PM coefficients represent conditional associations within a structural model. In contrast, SHAP values represent predictor contributions to fitted model predictions. Therefore, agreement between the two methods was interpreted as analytical consistency rather than independent theoretical validation. Moreover, interventional SHAP does not fully preserve dependencies between correlated predictors. This limitation may affect feature-importance attribution [50].

3.6. Decision-Support Prioritisation Logic

The final analysis examined how structural coefficients and predictive feature importance could be interpreted together. Its purpose was to evaluate the reliability of importance-based prioritisation. The analysis focused on socio-technical factors relevant to human–AI strategic decision support. Structural importance was measured using the absolute values of estimated PLS-PM coefficients. Predictive importance was measured using mean absolute SHAP values. Both measures were compared with the known population ranking presented in Table 1. The evaluation focused on two aspects. The first was identifying the strongest antecedent. The second was accurately recovering the complete ranking of all six predictors.

In the population model, PDAT had the largest absolute generating coefficient. Therefore, correctly identifying PDAT as the strongest antecedent provided a clear benchmark for ranking accuracy. However, identifying the highest-ranked predictor does not necessarily indicate that it is the most appropriate target for organisational intervention. Similarly, agreement between PLS-PM and SHAP rankings does not establish causal importance. It also does not justify changes to AI autonomy or delegated authority. Such managerial decisions require independent validation using real organisational data. They must also consider system reliability, institutional responsibilities, operational risks, and human oversight requirements. (see Table 5)

Figure 2 presents the complete computational workflow, from the prespecified MLAAR population model and synthetic data generation to PLS-PM estimation, bootstrap inference, cross-validated prediction, SHAP attribution, and performance evaluation. The workflow included 12 experimental conditions, with 100 replications per condition, yielding 1200 simulated datasets.

Table 5. Monte Carlo Performance Criteria for Structural, Predictive, and Prioritisation Reliability.

Evaluation Dimension	Performance Criterion	Interpretation
Structural recovery	Mean bias, relative bias, RMSE	Agreement with nominal generating coefficients
Directional recovery	All six coefficient signs correct (%)	Recovery of theoretical directions
Population-ranking recovery	Spearman correlation	Agreement with known predictor ordering
Strongest-predictor recovery	PDAT ranked first (%)	Identification of the strongest population antecedent
Structural–SHAP agreement	Identical complete rankings (%)	Consistency between estimated importance orderings
Direct and indirect associations	Estimation bias	Deviation from population values
Bootstrap inference	Detection and 95% CI coverage	Statistical detection and interval calibration
Predictive performance	Out-of-fold R^2	Predictive generalisation within the simulation
Predictive model comparison	OLS outperforming boosting (%)	Paired descriptive performance frequency

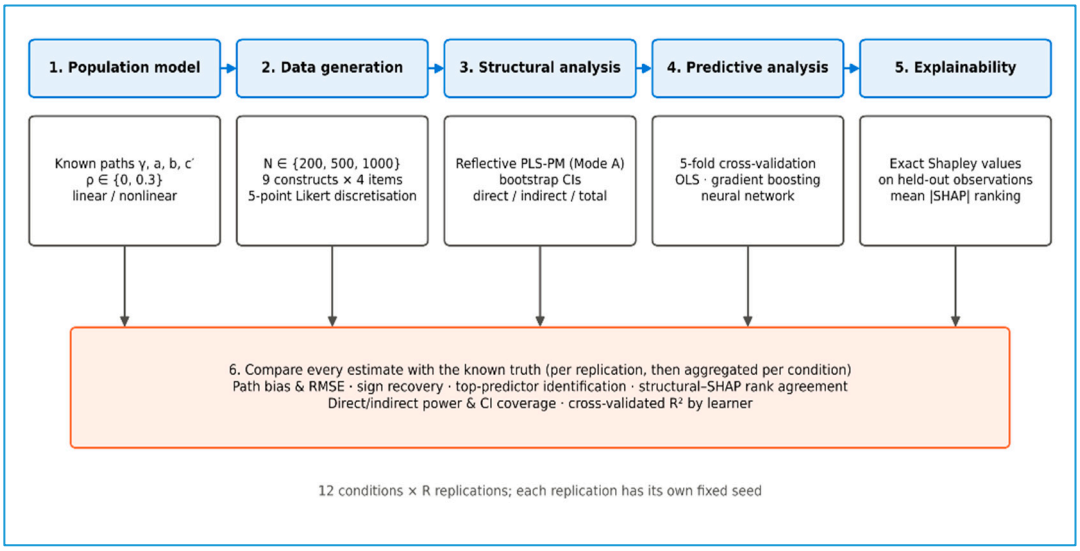


Figure 2. Monte Carlo simulation and hybrid structural–predictive analytical workflow.

The Monte Carlo simulation and hybrid structural–predictive analytical workflow illustrate six stages: population-model specification, synthetic data generation, structural estimation and bootstrap inference, cross-validated predictive modelling, held-out SHAP attribution, and evaluation against known population parameters.

The experiment comprises 12 conditions and 100 replications per condition, yielding 1200 simulated datasets. The computational procedures were implemented in Python 3.13 (Python Software Foundation, Wilmington, DE, USA) using NumPy 2.5.3, pandas 3.0.5, SciPy 1.18.1, scikit-learn 1.9.1 (including HistGradientBoost-ingRegressor and ML-PRegressor), and Matplotlib 3.11.2. The source code, exact soft-ware requirements, and documentation are available in the supplementary repository (https://doi.org/10.17605/OSF.IO/BZUYS). The study's decision-support contribution is methodological: it evaluates whether a hybrid structural and explainable predictive pipeline provides sufficiently

reliable importance information to warrant further consideration in empirical human–AI SDSS research.

3.7. Comparative Estimators and Reproducibility

To provide comparative benchmarks and support reproducibility, we reanalysed the same 1200 datasets using three alternative estimators. The original random seeds were retained. All estimators used identical bootstrap resamples within each replication (B = 100). The first estimator was consistent partial least squares (PLSc). It corrects attenuation in PLS-PM construct correlations using the reliability coefficient ρA [48]. The second was ordinary least squares (OLS) regression using unit-weight composites, calculated as the mean of each construct's indicators. The third was oracle regression using the true latent scores. Oracle regression is not feasible with ordinary empirical data. However, it provides a benchmark for distinguishing sampling error from measurement error.

PLSc is undefined when a construct's ρA falls outside (0, 1] or when the corrected correlation matrix is not positive-definite [48]. We therefore assessed PLSc admissibility separately for the antecedent and mediation equations. Only admissible solutions were included in the PLSc summaries. We evaluated coefficient recovery, bias, root mean squared error (RMSE), and bootstrap confidence-interval coverage. The same evaluation procedures were applied to all estimators. Appendix B (Table A2) documents the scripts, random seeds, software versions, computation times, and repository files.

4. Results

4.1. Data Characteristics and Measurement Quality

The Monte Carlo experiment generated 1200 synthetic datasets across 12 experimental conditions, with 100 replications per condition. Sample sizes were 200, 500, and 1000; antecedent intercorrelations were ρ = 0.0 and 0.3; and the generating relationships were linear or nonlinear. Each dataset contained nine latent constructs, each represented by four reflective indicators, yielding 36 five-category ordinal variables. All observations were computationally generated from the population model specified in Section 3; no participant or institutional records were analysed.

Measurement diagnostics were examined in greater detail using the separate illustrative dataset of 681 synthetic observations generated under the linear condition with ρ = 0.3. Table 6 presents the indicator-loading ranges, internal consistency, and convergent-validity results.

Table 6. Reflective Measurement Quality in the Illustrative Synthetic Dataset (N = 681).

Construct	Loading Range	Cronbach's α	CR
AIAA	0.765–0.880	0.834	0.889
AIDO	0.764–0.864	0.840	0.892
PDAT	0.755–0.892	0.837	0.891
HAIT	0.767–0.854	0.842	0.892
AIGO	0.754–0.851	0.848	0.889
HCAIO	0.233–0.824	0.836	0.721
MLAAR	0.791–0.871	0.854	0.901
SDMSE	0.758–0.864	0.837	0.891
IR	0.786–0.878	0.854	0.901

As shown in Table 6, all constructs exceeded the conventional internal-consistency criterion of 0.70. However, the estimated PLS-PM outer loadings for Human-Centric AI Orientation (HCAIO) ranged from 0.233 to 0.824, giving an AVE of 0.423, below the convergent-validity criterion of 0.50. Because all nine constructs were generated with the same population loadings (0.72–0.88), this result was examined with the diagnostics described in Section 3.2 (see Table 7).

Table 7. Measurement Diagnostics for the Illustrative Synthetic Dataset (N = 681).

Construct	Item–Own Latent *r*	Max. Cross *r*	Unit-Weight Loadings (AVE)	PLS-PM Loadings (AVE)	*r* (PLS Score, True Latent)	*r* with MLAAR: Sample (Population)
AIAA	0.673–0.827	0.286	0.668	0.666	0.918	0.196 (0.285)
AIDO	0.687–0.833	0.255	0.676	0.675	0.916	0.225 (0.299)
PDAT	0.677–0.831	0.339	0.673	0.672	0.921	0.301 (0.376)
HAIT	0.675–0.827	0.260	0.679	0.675	0.915	−0.069 (−0.009)
AIGO	0.693–0.840	0.248	0.687	0.668	0.892	0.042 (0.054)
HCAIO	0.683–0.810	0.264	0.670	0.423	0.710	0.015 (−0.030)
MLAAR	0.701–0.841	0.393	0.696	0.696	0.929	–
SDMSE	0.677–0.828	0.475	0.672	0.672	0.923	–
IR	0.697–0.838	0.466	0.695	0.695	0.930	–

The diagnostics indicate that the HCAIO result arises from PLS-PM estimation rather than from the measurement model or its implementation. Regenerating the illustrative dataset from its seed reproduced the published item file exactly. Each HCAIO indicator correlated 0.683–0.810 with the true HCAIO latent variable, within the range observed for the other constructs (0.673–0.841), and no indicator correlated more strongly with another construct's latent variable than with its own. Unit-weight composite loadings for HCAIO were 0.791–0.846 (AVE = 0.670; CR = 0.890), matching the other constructs.

In contrast, the PLS-PM outer weights for HCAIO were irregular, with a negative relative weight on the second indicator (−0.292), and the resulting construct score correlated only 0.710 with the true HCAIO latent variable, compared with 0.892–0.930 for the other constructs. The cause is the inner-model signal: with ρ = 0.3, the population correlation of HCAIO with MLAAR is −0.030, and its sample correlation in this dataset was 0.015. Its Mode-A outer weights were therefore driven by sampling error rather than by the common factor.

Across the 600 linear-condition replications, this pattern was systematic rather than specific to HCAIO or to this dataset (Appendix A, Table A1 and Figure A1). Under ρ = 0.3, the three antecedents whose population correlation with MLAAR is close to zero—HAIT (−0.009), AIGO (0.054), and HCAIO (−0.030)—had PLS-PM AVE values below 0.50 in 11–34% of replications, and increasing the sample size did not remove the problem, because the population signal itself is near zero.

AIAA, AIDO, and PDAT never fell below 0.50. Under $\rho = 0.0$, the same occurred mainly for AIGO, the antecedent with the weakest coefficient ($|\gamma| = 0.09$; 17% at N = 200 and 1% at N = 1000). No construct fell below 0.50 under unit weighting in any replication. The low HCAIO loading and AVE therefore reflect the interaction between the antecedent design and Mode-A outer weighting, not an error in indicator coding, mapping, or implementation.

Discriminant-validity and collinearity diagnostics were comparatively favourable. The maximum heterotrait–monotrait ratio (HTMT) was 0.549, while variance inflation factors

(VIFs) between the six MLAAR antecedents ranged from 1.104 to 1.169. The maximum absolute correlation between any two antecedent latent scores was approximately 0.250.

These results indicated no substantial overlap or multicollinearity under the selected screening criteria, including AIAA, AIDO, and PDAT. The HCAIO result discussed above therefore concerns PLS-PM outer weighting rather than overlapping or collinearity among the antecedents. Complete diagnostics are provided in the supplementary measurement, HTMT, latent-correlation, and VIF files.

4.2. Structural Coefficient Recovery in the Illustrative Dataset

Structural estimation assessed recovery of the prespecified MLAAR antecedent pathways (H1–H6) and downstream relationships involving Strategic Decision-Making System Effectiveness (SDMSE) and Institutional Resilience (IR; H7–H9). Table 8 presents the estimated coefficients from the illustrative synthetic dataset alongside their nominal population parameters and bootstrap confidence intervals.

Table 8. Illustrative Structural Coefficient Recovery and Bootstrap Confidence Intervals.

Structural Relationship	Population Value	Estimated β	95% Bootstrap CI
AIAA → MLAAR	0.240	0.157	[0.092, 0.240]
AIDO → MLAAR	0.260	0.179	[0.116, 0.253]
PDAT → MLAAR	0.370	0.282	[0.212, 0.356]
HAIT → MLAAR	−0.180	−0.206	[−0.279, −0.067]
AIGO → MLAAR	−0.090	−0.020	[−0.146, 0.085]
HCAIO → MLAAR	−0.210	−0.026	[−0.208, 0.085]
MLAAR → SDMSE	−0.400	−0.415	[−0.475, −0.353]
SDMSE → IR	0.450	0.433	[0.370, 0.495]
MLAAR → IR (direct)	−0.100	−0.090	[−0.162, −0.016]
MLAAR → SDMSE → IR (indirect)	−0.180	−0.180	[−0.219, −0.144]
MLAAR → IR (total)	−0.280	−0.270	[−0.339, −0.199]

PDAT showed the largest positive estimated association with MLAAR (β = 0.282), followed by AIDO (β = 0.179) and AIAA (β = 0.157). The corresponding negative estimates were −0.206 for HAIT, −0.020 for AIGO, and −0.026 for HCAIO. The MLAAR–SDMSE coefficient was −0.415, while the adjusted SDMSE–IR coefficient was 0.433. The estimated direct MLAAR–IR association was −0.090, and the indirect and total associations were approximately −0.180 and −0.270, respectively. These findings illustrate recovery within one generated dataset; they do not establish the estimator's reliability across repeated samples.

The PDAT–AIAA contrast was 0.125 (95% CI [0.004, 0.229], p = 0.042). However, PDAT–AIDO (0.103, 95% CI [−0.011, 0.209], p = 0.074) and AIDO–AIAA (0.023, 95% CI [−0.085, 0.115], p = 0.742) were not statistically distinguishable at the 5% level.

4.3. Within-Sample Robustness and Contextual Heterogeneity

Robustness was examined across the 12 Monte Carlo conditions. The analysis assessed whether structural recovery and bootstrap inference remained stable as sample size, antecedent correlation, and the generating functional form changed.

As shown in Table 9, at N = 1000 the mean absolute bias was 0.044 and the mean RMSE was 0.058, and all six coefficient signs were recovered in 95.3% of replications. Relative bias, reported as a percentage of the population coefficient, ranged from −13.2% to −12.9%

under ρ = 0.0 and from −28.0% to −27.2% under ρ = 0.3 at this sample size. The bootstrap results revealed contrasting patterns for direct and indirect MLAAR–IR associations.

Table 9. Structural Coefficient Recovery Across Monte Carlo Conditions.

N	ρ	Form	Mean |Bias|	Relative Bias (%)	RMSE	All Signs Correct (%)
200	0.0	Linear	0.027	−10.37	0.068	84
200	0.0	Nonlinear	0.030	−9.57	0.068	92
200	0.3	Linear	0.059	−30.61	0.105	41
200	0.3	Nonlinear	0.067	−35.49	0.110	38
500	0.0	Linear	0.029	−10.90	0.048	99
500	0.0	Nonlinear	0.031	−12.91	0.051	96
500	0.3	Linear	0.062	−31.03	0.086	69
500	0.3	Nonlinear	0.058	−27.55	0.083	78
1000	0.0	Linear	0.031	−13.24	0.043	99
1000	0.0	Nonlinear	0.032	−12.89	0.045	100
1000	0.3	Linear	0.056	−27.15	0.073	88
1000	0.3	Nonlinear	0.057	−28.00	0.070	94

Direct-effect bias remained comparatively small, ranging from −0.018 to 0.002, and average direct-effect confidence-interval coverage was approximately 93.3% (MCSE = 0.7 percentage points). Direct-effect detection increased from 44.0% (MCSE = 2.5) at N = 200 to 95.5% (MCSE = 1.0) at N = 1000. The indirect association exhibited persistent positive bias, averaging approximately +0.049 (range 0.043–0.053), meaning the estimates were attenuated towards zero.

Indirect-effect detection reached 100% in all conditions, but mean 95% confidence-interval coverage declined from 65.8% (MCSE = 2.4) at N = 200 to 34.0% (MCSE = 2.4) at N = 500 and 9.8% (MCSE = 1.5) at N = 1000. Across all replications, coverage averaged approximately 36.5% (MCSE = 1.4). Monte Carlo standard errors for every condition are reported in Table 10; pooled values are based on 400 replications per sample size. (see Table 10)

Figure 3 illustrates the increase in direct-effect detection and the simultaneous decline in confidence-interval coverage for indirect effects. The contrast demonstrates why detection probability and interval calibration must be evaluated separately.

The bootstrap sensitivity analysis showed that these conclusions do not depend on the number of resamples (Table 11). Increasing the budget from B = 100 to B = 1000 on the same datasets changed indirect-effect coverage by at most 5 percentage points in any condition, direct-effect coverage by at most 2 points, and direct-effect detection by at most 5 points, all within approximately one Monte Carlo standard error.

Indirect-effect coverage with B = 1000 remained 68–75% at N = 200, 35% at N = 500, and 12–14% at N = 1000, and indirect-effect detection remained 100%. The larger budget widened the intervals only slightly (mean indirect-interval width at N = 1000: 0.055 with B = 100 and 0.057 with B = 1000). The undercoverage of the indirect effect therefore reflects estimator bias rather than an insufficient bootstrap budget.

Across the six conditions, the mean indirect estimate was approximately −0.13 against the population value of −0.18, a bias of about +0.05 that did not shrink with sample size. Relative to the mean half-width of the B = 1000 intervals, this bias was 0.73 at N = 200,

1.28 at N = 500, and 1.72 at N = 1000. As the intervals narrow with increasing N, they increasingly exclude the population value, regardless of how many resamples are used.

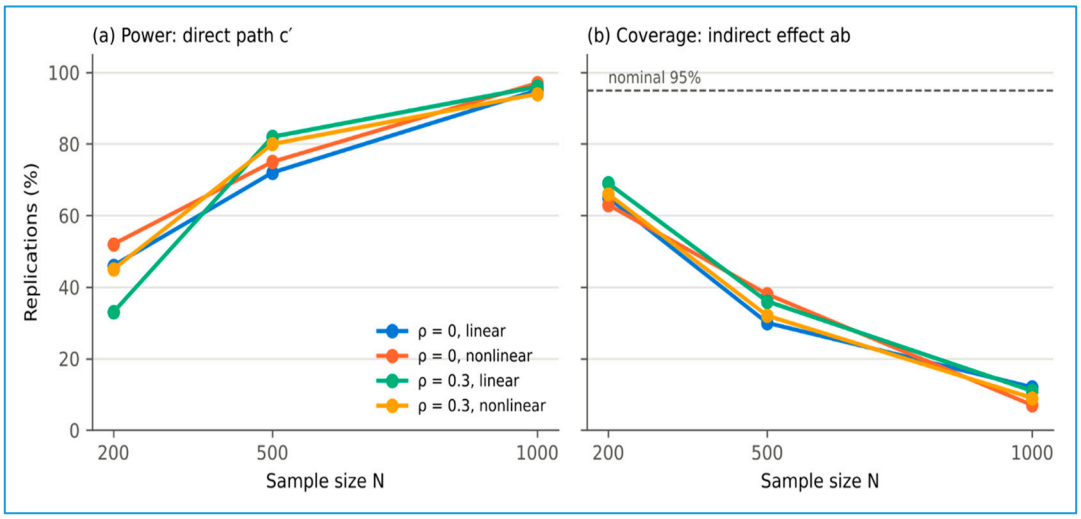


Figure 3. Bootstrap detection and coverage.

Table 10. Bootstrap Detection and 95% Confidence-Interval Coverage for the Direct and Indirect MLAAR–IR Associations Across Monte Carlo Conditions (B = 100 Resamples per Replication).

N	ρ	Form	Direct Detection (%)	Direct Coverage (%)	Indirect Detection (%)	Indirect Coverage (%)
200	0.0	Linear	46 (5.0)	95 (2.2)	100 (0.0)	65 (4.8)
200	0.0	Nonlinear	52 (5.0)	95 (2.2)	100 (0.0)	63 (4.8)
200	0.3	Linear	33 (4.7)	97 (1.7)	100 (0.0)	69 (4.6)
200	0.3	Nonlinear	45 (5.0)	93 (2.6)	100 (0.0)	66 (4.7)
500	0.0	Linear	72 (4.5)	94 (2.4)	100 (0.0)	30 (4.6)
500	0.0	Nonlinear	75 (4.3)	94 (2.4)	100 (0.0)	38 (4.9)
500	0.3	Linear	82 (3.8)	87 (3.4)	100 (0.0)	36 (4.8)
500	0.3	Nonlinear	80 (4.0)	94 (2.4)	100 (0.0)	32 (4.7)
1000	0.0	Linear	95 (2.2)	93 (2.6)	100 (0.0)	12 (3.2)
1000	0.0	Nonlinear	97 (1.7)	96 (2.0)	100 (0.0)	7 (2.6)
1000	0.3	Linear	96 (2.0)	90 (3.0)	100 (0.0)	11 (3.1)
1000	0.3	Nonlinear	94 (2.4)	91 (2.9)	100 (0.0)	9 (2.9)

Table 11. Bootstrap Sensitivity Analysis: Detection and Coverage with 100 versus 1000 Resamples on Identical Datasets.

N	ρ	B	Direct Detection (%)	Direct Coverage (%)	Indirect Detection (%)	Indirect Coverage (%)	Mean Indirect CI Width
200	0.0	100	45 (5.0)	95 (2.2)	100 (0.0)	71 (4.5)	0.121
200	0.0	1000	44 (5.0)	96 (2.0)	100 (0.0)	68 (4.7)	0.126
500	0.0	100	73 (4.4)	97 (1.7)	100 (0.0)	33 (4.7)	0.078
500	0.0	1000	72 (4.5)	97 (1.7)	100 (0.0)	35 (4.8)	0.081

Table 11. *Cont.*

N	ρ	B	Direct Detection (%)	Direct Coverage (%)	Indirect Detection (%)	Indirect Coverage (%)	Mean Indirect CI Width
1000	0.0	100	96 (2.0)	91 (2.9)	100 (0.0)	9 (2.9)	0.054
1000	0.0	1000	95 (2.2)	93 (2.6)	100 (0.0)	12 (3.2)	0.057
200	0.3	100	38 (4.9)	95 (2.2)	100 (0.0)	73 (4.4)	0.125
200	0.3	1000	33 (4.7)	96 (2.0)	100 (0.0)	75 (4.3)	0.131
500	0.3	100	85 (3.6)	88 (3.2)	100 (0.0)	30 (4.6)	0.078
500	0.3	1000	81 (3.9)	89 (3.1)	100 (0.0)	35 (4.8)	0.080
1000	0.3	100	93 (2.6)	92 (2.7)	100 (0.0)	15 (3.6)	0.055
1000	0.3	1000	95 (2.2)	92 (2.7)	100 (0.0)	14 (3.5)	0.057

4.4. Out-of-Sample Predictive Performance

Predictive performance was evaluated using five-fold cross-validation. Ordinary least squares (OLS) regression, gradient boosting, and an artificial neural network (ANN) were compared using identical held-out partitions. Table 12 reports the mean cross-validated coefficient of determination across the 12 experimental conditions.

Table 12. Cross-Validated Predictive Performance Across Monte Carlo Conditions.

N	ρ	Form	OLS R^2	GBM R^2	ANN R^2	OLS > GBM (%)
200	0.0	Linear	0.209	0.089	0.053	98
200	0.0	Nonlinear	0.195	0.103	0.068	93
200	0.3	Linear	0.157	0.035	−0.007	100
200	0.3	Nonlinear	0.151	0.064	0.043	92
500	0.0	Linear	0.231	0.157	0.144	100
500	0.0	Nonlinear	0.225	0.186	0.175	91
500	0.3	Linear	0.180	0.112	0.095	100
500	0.3	Nonlinear	0.178	0.147	0.133	85
1000	0.0	Linear	0.241	0.198	0.196	100
1000	0.0	Nonlinear	0.239	0.227	0.227	77
1000	0.3	Linear	0.192	0.151	0.143	100
1000	0.3	Nonlinear	0.182	0.176	0.172	63

OLS achieved the highest mean cross-validated R^2 in all 12 conditions. Across the experiment, mean R^2 values were approximately 0.198 for OLS, 0.137 for gradient boosting, and 0.120 for the ANN. OLS outperformed gradient boosting in approximately 91.6% of replications. The performance differences between OLS and the nonlinear learners also narrowed as sample size increased, particularly under nonlinear generating conditions. At N = 1000 and ρ = 0.3 under nonlinear generation, OLS achieved $R^2 = 0.182$, compared with 0.176 for gradient boosting and 0.172 for the ANN. Regarding this condition, OLS outperformed gradient boosting in 63% of replications.

Figure 4 compares mean cross-validated R^2 across sample sizes for linear and nonlinear generating processes. The separate illustrative dataset produced mean R^2 values of 0.146 for OLS, 0.095 for gradient boosting, and 0.110 for the ANN across 25 held-out evaluations. Corresponding RMSE values were 0.696, 0.716, and 0.710, respectively. OLS outperformed gradient boosting in 22 of the 25 evaluations.

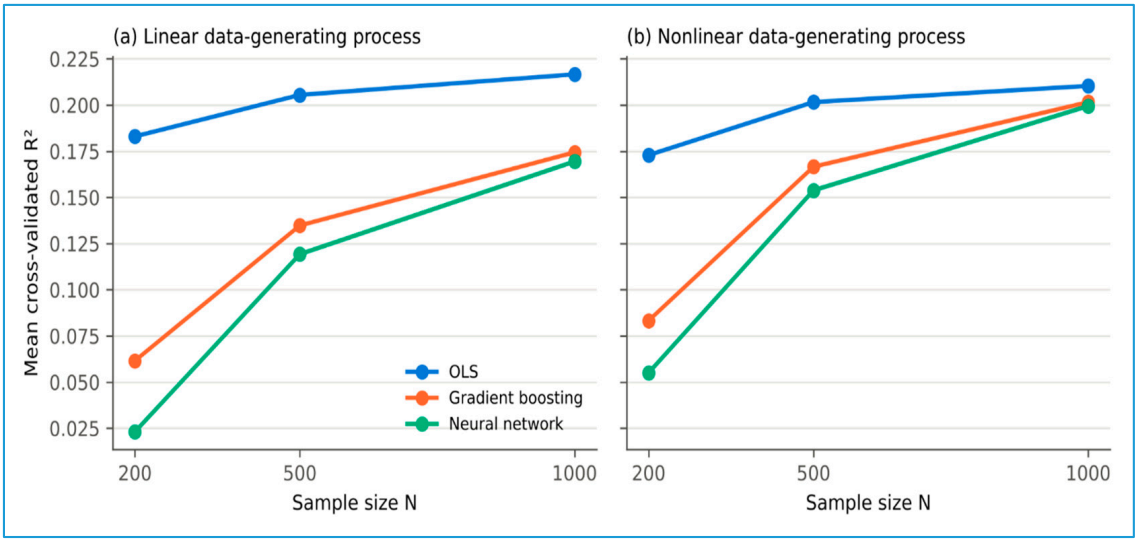


Figure 4. Mean cross-validated R^2 by sample size and algorithm: (**a**) linear and (**b**) nonlinear conditions, averaged over ρ.

4.5. Explainable Machine Learning Results

SHapley Additive exPlanations (SHAP) were calculated for held-out predictions from the gradient-boosting models. The analysis assessed whether SHAP importance recovered the known population ordering and correctly identified PDAT as the strongest antecedent. The resulting rankings were also compared with those obtained from the absolute PLS-PM structural coefficients.

As shown in Table 13, PDAT was ranked first by PLS-PM in 70.0%, 88.3%, and 96.3% of replications at N = 200, 500, and 1000, respectively. The corresponding SHAP percentages were 71.0%, 88.3%, and 93.3%. The mean Spearman correlations were approximately 0.801 for PLS-PM and 0.804 for SHAP. Exact agreement between the two estimated rankings occurred in only 22.3% of the 1200 replications, ranging from 5% to 44% across conditions.

Table 13. Predictor-Ranking Recovery and Structural–SHAP Agreement Across Monte Carlo Conditions.

N	ρ	Form	PLS Top PDAT (%)	SHAP Top PDAT (%)	Population–PLS ρ_s	Population–SHAP ρ_s	Exact Agreement (%)
200	0.0	Linear	73	68	0.719	0.678	12
200	0.0	Nonlinear	81	81	0.689	0.665	17
200	0.3	Linear	63	62	0.685	0.689	5
200	0.3	Nonlinear	63	73	0.630	0.684	5
500	0.0	Linear	89	86	0.862	0.817	29
500	0.0	Nonlinear	91	90	0.861	0.847	40
500	0.3	Linear	87	86	0.795	0.853	14
500	0.3	Nonlinear	86	91	0.792	0.837	20
1000	0.0	Linear	99	95	0.927	0.888	39
1000	0.0	Nonlinear	98	98	0.901	0.879	44
1000	0.3	Linear	93	86	0.875	0.897	21
1000	0.3	Nonlinear	95	94	0.875	0.908	21

Figure 5 illustrates the differences between recovery of the strongest predictor and complete ranking agreement. Although both methods increasingly identified PDAT cor-

rectly, their complete rankings frequently differed. This result shows that agreement on the highest-ranked predictor does not imply full explanatory–predictive convergence.

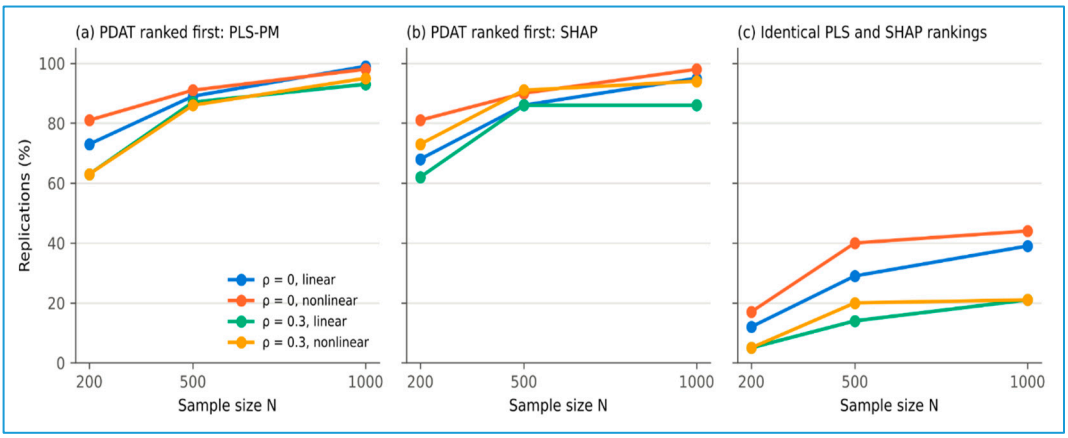


Figure 5. Predictor-ranking recovery.

4.6. Structural–Predictive Alignment and Managerial Interpretation

The final analysis examined whether structural and predictive importance measures provided consistent information when assessed against known generating relationships.

Figure 6 illustrates a comparison of absolute population coefficients, estimated PLS-PM coefficients, and held-out SHAP importance values. However, the complete rankings differed. Spearman's correlation with the population ordering was 0.657 for PLS-PM and 0.943 for SHAP, while agreement between the two estimated rankings was 0.829.

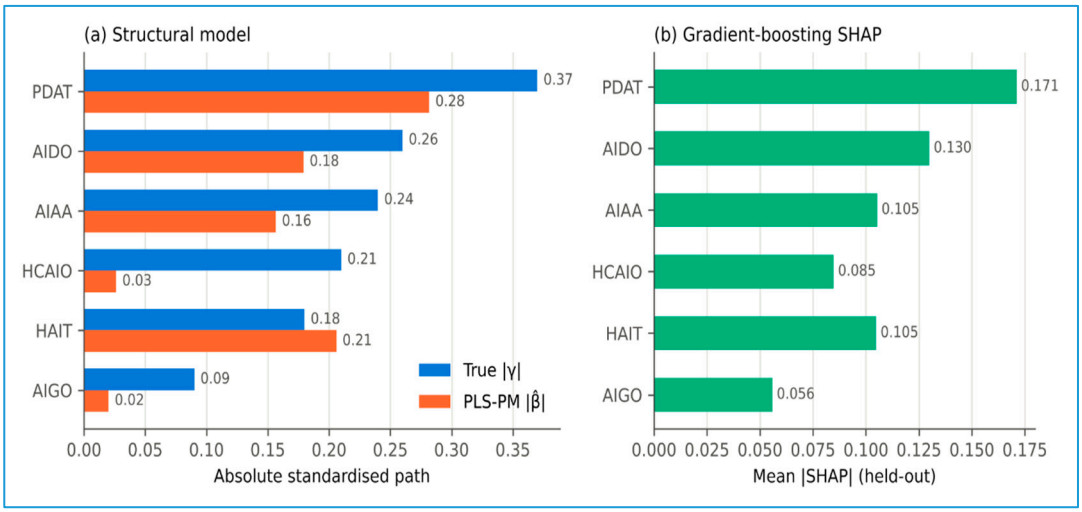


Figure 6. Illustrative structural–predictive alignment.

Thus, the SHAP importance ranking more accurately reproduced the known population ordering in this individual dataset. However, the PLS-PM and SHAP estimates also showed relatively strong agreement with each other. The difference was particularly noticeable for HCAIO. Its absolute population coefficient was 0.210, whereas its estimated absolute PLS-PM coefficient was only 0.026.

4.7. Comparative Analysis of Alternative Estimators

Table 14 compares four estimators using the same synthetic datasets. Results were pooled across the two functional forms, with 200 replications per condition. This comparison helps explain the estimation bias reported in Sections 4.3 and 4.5. For the indirect

effect, oracle regression produced negligible bias (−0.001 to 0.001). Its confidence-interval coverage ranged from 92.5% to 94.0%. These results indicate that the percentile bootstrap performed adequately when the true latent scores were available.

Table 14. Comparison of PLS-PM with Consistent PLS, Unit-Weight Composites, and an Oracle Regression on the Same Simulated Datasets.

ρ	N	Estimator	Admissible (%)	Relative Bias (%)	RMSE	All Signs Correct (%)	Indirect Bias	Indirect Coverage (%)
0.0	200	PLS-PM	100.0	−10.0	0.068	88.0	0.049	67.5 (3.3)
0.0	200	PLSc	53.0	9.3	0.073	95.3	−0.001	93.0 (1.8)
0.0	200	Unit-weight	100.0	−15.3	0.070	92.0	0.053	61.0 (3.4)
0.0	200	Oracle	100.0	1.7	0.056	95.5	−0.001	94.0 (1.7)
0.0	500	PLS-PM	100.0	−11.9	0.050	97.5	0.050	34.0 (3.3)
0.0	500	PLSc	80.0	4.9	0.051	99.4	−0.001	94.0 (1.7)
0.0	500	Unit-weight	100.0	−15.2	0.053	97.0	0.053	26.5 (3.1)
0.0	500	Oracle	100.0	0.8	0.035	99.5	0.001	93.0 (1.8)
0.0	1000	PLS-PM	100.0	−13.1	0.044	99.5	0.051	9.0 (2.0)
0.0	1000	PLSc	92.5	2.0	0.031	100.0	0.000	91.5 (2.0)
0.0	1000	Unit-weight	100.0	−14.7	0.046	100.0	0.053	7.5 (1.9)
0.0	1000	Oracle	100.0	0.7	0.026	100.0	0.000	93.5 (1.7)
0.3	200	PLS-PM	100.0	−33.1	0.107	39.5	0.045	70.0 (3.2)
0.3	200	PLSc	17.0	−15.0	0.130	47.1	−0.007	89.5 (2.2)
0.3	200	Unit-weight	100.0	−25.6	0.089	72.5	0.050	61.5 (3.4)
0.3	200	Oracle	100.0	−1.7	0.068	90.5	0.000	92.5 (1.9)
0.3	500	PLS-PM	100.0	−29.3	0.084	73.5	0.051	33.5 (3.3)
0.3	500	PLSc	19.5	−1.4	0.127	76.9	0.000	92.5 (1.9)
0.3	500	Unit-weight	100.0	−22.6	0.065	95.0	0.053	29.0 (3.2)
0.3	500	Oracle	100.0	−1.1	0.042	98.0	0.001	92.5 (1.9)
0.3	1000	PLS-PM	100.0	−27.6	0.071	91.0	0.050	7.5 (1.9)
0.3	1000	PLSc	28.0	1.5	0.079	91.1	−0.001	90.0 (2.1)
0.3	1000	Unit-weight	100.0	−22.6	0.058	99.0	0.052	7.0 (1.8)
0.3	1000	Oracle	100.0	0.5	0.030	100.0	0.000	93.0 (1.8)

Consistent PLS (PLSc) also substantially reduced bias (−0.007 to 0.000). Its mediation solutions were admissible in all 1200 replications. Coverage improved to 89.5–94.0%. In contrast, PLS-PM and unit-weight composites retained a positive bias of approximately 0.05. Their coverage declined from 61–70% at N = 200 to 7–9% at N = 1000. These findings indicate that measurement attenuation was the main source of undercoverage in indirect-effect estimates under the simulated conditions.

Although PLSc largely corrected this problem, its wider confidence intervals reduced direct-effect detection to 25.0–81.0%, compared with 37.0–94.5% for PLS-PM. The antecedent-path results revealed a different trade-off. PLSc reduced relative bias to −15.0% to 9.3%, compared with −33.1% to −10.0% for PLS-PM. However, PLSc frequently produced inadmissible solutions. Its antecedent model was admissible in 53.0–92.5% of replications when ρ = 0.0. This proportion declined to 17.0–28.0% when ρ = 0.3. HAIT, AIGO, and HCAIO accounted for 97.1% of block-level failures. These were the weakly

connected antecedents identified in Section 4.1. Even among admissible solutions, PLSc had a higher RMSE than PLS-PM when ρ = 0.3 (0.079–0.130 versus 0.071–0.107).

Unit-weight composites showed slightly greater attenuation than PLS-PM when ρ = 0.0. Their relative bias ranged from −15.3% to −14.7%, compared with −13.1% to −10.0% for PLS-PM. However, unit-weight composites showed less attenuation when ρ = 0.3 (−25.6% to −22.6% versus −33.1% to −27.6%). They also achieved lower RMSE and more accurate directional recovery under correlated conditions. All six coefficient signs were recovered in 72.5–99.0% of replications, compared with 39.5–91.0% for PLS-PM.

Oracle regression provided the benchmark without measurement error. Its relative bias ranged from −1.7% to 1.7%. It also recovered all six coefficient signs in 90.5–100% of replications. Overall, the comparison shows that no feasible estimator performed best across all criteria. PLSc improved bias and indirect-effect coverage but faced substantial admissibility limitations. Unit-weight composites offered more stable estimation under correlated conditions. These findings highlight the importance of evaluating bias, precision, admissibility, and confidence-interval coverage separately. (see Table 14)

5. Discussion

This study evaluated the methodological reliability of a hybrid framework for human–AI Strategic Decision Support Systems (SDSSs). The framework integrated partial least squares path modelling (PLS-PM), predictive machine learning, and SHapley Additive exPlanations (SHAP). The measurement analysis revealed important limitations. Acceptable internal consistency, discriminant validity, and low multicollinearity did not guarantee adequate convergent validity in the estimates. This limitation was particularly evident for Human-Centric AI Orientation (HCAIO). Its estimated PLS-PM outer loadings and average variance extracted (AVE) were low, although all constructs shared the same population measurement model.

Additional diagnostics linked this result to Mode-A outer weighting. When an exogenous construct had a weak correlation with its adjacent endogenous construct, sampling error strongly influenced its outer weight. This reduced the accuracy of its estimated construct score and attenuated its structural coefficient. For HCAIO, the estimated coefficient was −0.026, compared with the population value of −0.210. Similar patterns affected Human–AI Trust (HAIT) and AI Governance and Human Oversight (AIGO) across simulation replications.

These findings indicate that low PLS-PM AVE does not necessarily reflect poor indicator quality. Structural-model-independent diagnostics can help distinguish weighting artefacts from measurement problems. Such diagnostics include unit-weight composite loadings and factor-based measurement estimates. Researchers should therefore assess individual measurement blocks alongside overall model diagnostics.

These findings are consistent with Dijkstra and Henseler [48], who identified inconsistencies in conventional PLS path estimates under reflective measurement. They also agree with Schuberth et al. [49], who highlighted bias in composite-based parameter estimation. However, this study extends the literature by examining these limitations within an integrated PLS-PM–machine learning–SHAP framework. It also considers ordinal indicators, correlated antecedents, and known population-ranking benchmarks.

The mediation analysis revealed an important difference between statistical detection and confidence-interval coverage. The direct association between MLAAR and Institutional Resilience (IR) showed relatively small estimation bias. It also achieved generally satisfactory confidence-interval coverage. In contrast, the indirect association through Strategic

Decision-Making System Effectiveness (SDMSE) showed persistent estimation bias. Its confidence intervals also exhibited inadequate coverage.

Although the indirect association was consistently detected, its confidence intervals frequently excluded the true population value. This pattern may reflect differences between component-based estimates and the underlying latent population parameters. A sensitivity analysis using 1000 bootstrap resamples confirmed that the bootstrap budget did not primarily cause the undercoverage. Instead, it resulted from estimator bias. At N = 500 and N = 1000, the bias exceeded half the confidence interval width. This led to inadequate coverage of the population parameter. These findings align with Morris et al. [46], who emphasised the importance of evaluating bias, precision, and confidence-interval coverage separately in simulation studies. They also relate to Yuan et al. [53], who highlighted how estimator selection, measurement error, and parameter scaling affect mediation analysis.

However, the present study differs from Yuan et al. It evaluates estimation bias and confidence-interval coverage against known population parameters. In contrast, Yuan et al. primarily compared mediation estimators using scale-independent signal-to-noise ratios. The predictive analysis showed that ordinary least squares (OLS) regression achieved the best average out-of-fold performance between the evaluated algorithms. However, the performance gap between OLS and nonlinear models narrowed as sample size increased. This pattern was particularly evident under nonlinear generating conditions.

These findings suggest that greater model complexity does not always improve predictive generalisation. Measurement error, limited sample sizes, and differences in estimation requirements may affect model performance. The nonlinear simulation conditions included prespecified interaction and quadratic effects. Nevertheless, the linear benchmark remained competitive. Therefore, interpret predictive performance in relation to the characteristics of the generating process. These include signal strength, functional form, sample size, preprocessing procedures, and algorithm configuration. Predictive comparisons should not be treated as direct tests of structural theory. These findings are consistent with comparative machine-learning research [55]. Such research indicates that predictive performance depends on dataset characteristics. It also shows that more complex algorithms do not consistently outperform simpler models.

PLS-PM coefficients represent conditional associations within a specified structural model. In contrast, SHAP values measure predictor contributions to predictions generated by a fitted machine-learning model. Therefore, their rankings may differ because of antecedent correlations, measurement error, and nonlinear relationships. Interventional SHAP also has limitations when predictors are correlated. It uses a reference distribution that does not fully preserve the observed dependencies between predictors. These findings are consistent with Hu et al. [50]. Their study highlighted the limitations of conventional Shapley explanations when predictors are correlated.

This study extends this discussion by comparing PLS-PM and SHAP rankings with a known population ordering. It also examines their agreement under controlled simulation conditions. This approach distinguishes accurate recovery of the population ranking from agreement between analytical methods. Consequently, agreement between PLS-PM and SHAP rankings does not independently establish causal importance or validate the theoretical MLAAR framework. This distinction has important methodological implications for interpretable human–AI SDSSs. The MLAAR framework demonstrates these challenges through six theoretically specified socio-technical antecedents. It also considers their downstream relationships with Strategic Decision-Making System Effectiveness (SDMSE) and Institutional Resilience (IR).

6. Theoretical, Methodological, and Practical Implications

Theoretically, this study contributes to research on human–AI Strategic Decision Support Systems (SDSSs). It introduces a theory-derived framework for examining relationships among technological, human, and organisational factors. The Millennial Leader–AI Agent Resistance (MLAAR) model distinguishes three key factors: AI Agent Autonomy (AIAA), AI Decision Opacity (AIDO), and Perceived Decision-Authority Threat (PDAT). It also incorporates trust, governance, and human-centric orientation. The framework extends to organisational outcomes through its proposed relationships with Strategic Decision-Making System Effectiveness (SDMSE) and Institutional Resilience (IR).

Methodologically, the study develops a reproducible Monte Carlo simulation framework. It evaluates the reliability of an integrated analytical approach combining partial least squares path modelling (PLS-PM), predictive machine learning, and SHapley Additive exPlanations (SHAP). The findings reveal important limitations. Correct coefficient directions, strong predictive accuracy, and agreement between predictor rankings do not necessarily indicate accurate recovery of population relationships. Persistent coefficient bias and inadequate confidence-interval coverage for indirect effects highlight the need for separate evaluations of estimation accuracy and inferential reliability.

Furthermore, superior linear predictive performance does not exclude nonlinear relationships in the generating process. Similarly, agreement between PLS-PM and SHAP rankings does not independently validate the theoretical model. Practically, the study provides methodological guidance for evaluating analytical evidence in human-centred AI governance. It highlights the importance of validating structural estimates, predictive models, and explainability results before using them to inform organisational decision-making.

7. Limitations and Future Research

The Monte Carlo simulation examined three sample sizes, two antecedent-correlation conditions, and two generating functional forms (linear and nonlinear). These conditions allowed for a systematic methodological evaluation. However, they do not capture the full complexity of human–AI strategic decision-making. Future simulations should examine more diverse correlation structures, different coefficient magnitudes, missing data, and measurement non-invariance. They should also consider more complex nonlinear relationships. The study estimated structural coefficients using reflective Mode-A partial least squares path modelling (PLS-PM). These estimates were compared with known latent population coefficients. Persistent estimation bias may result from measurement error, ordinal discretisation, or differences between the estimator's target and the generating parameters. Section 4.7 compared PLS-PM with consistent PLS, unit-weight composites, and an oracle regression. Future research should extend this comparison to covariance-based structural equation modelling with ordinal (polychoric) estimation and to factor-score regression. The measurement diagnostics also revealed instability in Mode-A's outer weights. This instability was particularly evident for antecedents weakly correlated with their endogenous construct. Future research should examine whether alternative path weighting schemes, consistent partial least squares (PLSc), or factor-based estimation can reduce this instability.

The main simulation used 100 bootstrap resamples per replication. A sensitivity analysis increased this number to 1000 for the six linear conditions. Detection and coverage rates differed by approximately one Monte Carlo standard error or less. These findings indicate that the low indirect-effect coverage was primarily associated with estimator bias rather than the bootstrap budget. Nevertheless, future studies should evaluate bias-corrected and alternative confidence-interval procedures. They should also examine estimator-specific reference parameters for composite-based mediation estimates.

The predictive analysis included only three algorithms with fixed configurations. This restriction may limit the generalisability of predictive comparisons. In addition, correlations among predictors may affect interventional SHAP explanations. Future research should examine additional predictive algorithms and apply more extensive hyperparameter optimisation. It should also consider dependence-aware explainability methods and independent external validation. These extensions could improve the reliability and interpretation of integrated explanatory–predictive frameworks.

8. Conclusions

This study evaluated the methodological reliability of a hybrid analytical framework for human–AI Strategic Decision Support Systems (SDSSs). The framework combined structural modelling, predictive machine learning, and explainable artificial intelligence (AI). It used the theory-derived Millennial Leader–AI Agent Resistance (MLAAR) model as a simulation framework with known parameters. A reproducible Monte Carlo experiment examined structural coefficient recovery, mediation inference, predictive performance, and predictor-ranking accuracy across 1200 simulated datasets.

The findings showed that larger sample sizes generally improved estimation precision and the recovery of coefficient directions. In contrast, correlations among antecedents increased bias in structural estimation. Perceived Decision-Authority Threat (PDAT) was often the strongest prespecified population antecedent. However, PLS-PM and SHAP importance rankings rarely agreed completely. The mediation analysis also revealed important limitations. Indirect associations were consistently detected, but their estimates showed systematic bias and inadequate confidence-interval coverage.

Ordinary least squares (OLS) regression achieved the strongest average cross-validated predictive performance among the evaluated algorithms. However, this finding did not exclude nonlinear relationships in the generating process. It also did not establish the general superiority of linear models beyond the simulation conditions examined. Similarly, agreement between structural coefficients and SHAP rankings indicated analytical consistency rather than independent theoretical validation.

This study's main contribution is a transparent, reproducible computational framework. It enables researchers to evaluate structural modelling, predictive machine learning, and explainable AI separately and assess their reliability. The findings highlight the importance of assessing measurement quality, parameter recovery, inferential calibration, predictive generalisation, and feature attribution. These evaluations are essential before drawing substantive conclusions or applying analytical findings to organisational decision-making.

Funding: The Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Project No. KFU266241), funded this work.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: During the preparation of this study, the authors used Claude Opus 5.5 (Anthropic, San Francisco, CA, USA; accessed October 2026) to assist with programming the simulation and code and to draft and edit the manuscript. The authors designed the study and specified the population model. They verified all code, results and analysis, which are openly available and fully reproducible (see the Data Availability Statement). The authors reviewed and edited all AI-assisted content and take full responsibility for the integrity and accuracy of this publication.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

Abbreviation	Full Form
AI	Artificial Intelligence
AIAA	AI Agent Autonomy
AIDO	AI Decision Opacity
AIGO	AI Governance and Human Oversight
ANN	Artificial Neural Network
AVE	Average Variance Extracted
CB-SEM	Covariance-Based Structural Equation Modelling
CI	Confidence Interval
CR	Composite Reliability
CV	Cross-Validation (as in R^2CV)
GBM	Gradient-Boosting Machine
H1–H9	Hypotheses 1–9
HAIT	Human–AI Trust
HCAIO	Human-Centric AI Orientation
HTMT	Heterotrait–Monotrait Ratio
IE	Indirect Effect
IR	Institutional Resilience
L-BFGS	Limited-Memory Broyden–Fletcher–Goldfarb–Shanno (optimisation algorithm)
MAE	Mean Absolute Error
MCSE	Monte Carlo Standard Error
MLAAR	Millennial Leader–AI Agent Resistance
NL	Nonlinear Component of the Generating Model (Equation (3))
OLS	Ordinary Least Squares
OSF	Open Science Framework
PDAT	Perceived Decision-Authority Threat
PLS	Partial Least Squares
PLSc	Consistent Partial Least Squares
PLS-PM	Partial Least Squares Path Modelling
PLS-SEM	Partial Least Squares Structural Equation Modelling
R^2	Coefficient of Determination
RMSE	Root Mean Squared Error
RQ1–RQ5	Research Questions 1–5
SDMSE	Strategic Decision-Making System Effectiveness
SDSS(s)	Strategic Decision Support System(s)
SEM	Structural Equation Modelling
SEM–ANN	Two-Stage Structural Equation Modelling–Artificial Neural Network Approach
SHAP	SHapley Additive exPlanations
TE	Total Effect
VIF	Variance Inflation Factor
XAI	Explainable Artificial Intelligence

Appendix A

This appendix reports the measurement diagnostics summarised in Section 4.1 for the 600 linear-condition replications (100 per combination of N and ρ), using the same seeds as the main simulation. Table A1 gives, for each antecedent, the percentage of replications in which the AVE computed from PLS-PM outer loadings fell below 0.50, the mean correlation between the PLS-PM score and the true latent variable, and the corresponding percentage under unit weighting. Figure A1 plots each antecedent's PLS-PM and unit-weight AVE against the absolute sample correlation between the antecedent and MLAAR.

Table A1. PLS-PM and Unit-Weight Convergent Validity of the Six Antecedents Across 600 Linear-Condition Replications.

Construct	ρ	Population r with MLAAR	PLS AVE < 0.50 (%), N = 200	N = 500	N = 1000	r (PLS Score, True), N = 1000	Unit-Weight AVE < 0.50 (%)
AIAA	0.0	0.240	1	0	0	0.920	0
AIAA	0.3	0.285	0	0	0	0.920	0
AIDO	0.0	0.260	0	0	0	0.920	0
AIDO	0.3	0.299	0	0	0	0.921	0
PDAT	0.0	0.370	0	0	0	0.921	0
PDAT	0.3	0.376	0	0	0	0.920	0
HAIT	0.0	−0.180	3	0	0	0.919	0
HAIT	0.3	−0.009	27	29	34	0.735	0
AIGO	0.0	−0.090	17	4	1	0.906	0
AIGO	0.3	0.054	21	19	11	0.845	0
HCAIO	0.0	−0.210	3	0	0	0.920	0
HCAIO	0.3	−0.030	23	22	19	0.816	0

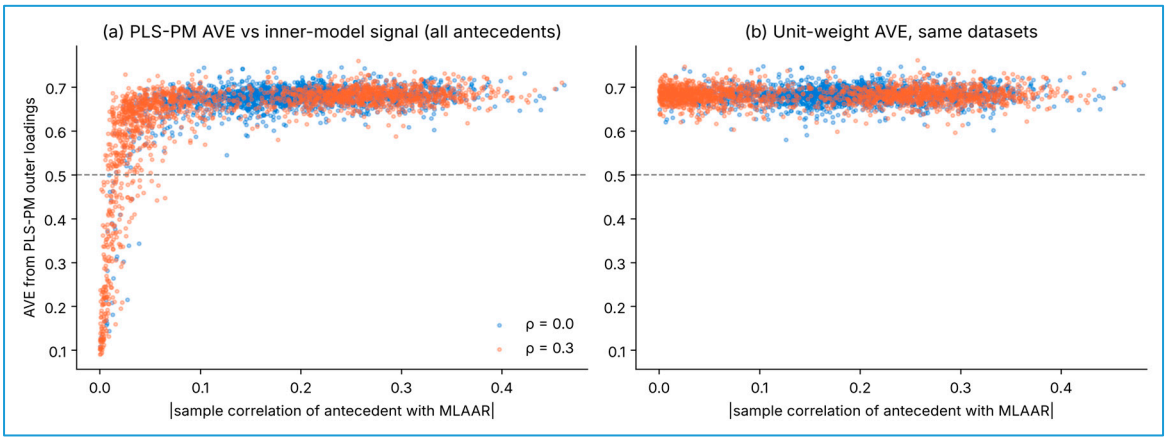


Figure A1. Average variance extracted of the six antecedents across 600 linear-condition replications: (**a**) from PLS-PM outer loadings and (**b**) from unit-weight loadings, against the absolute sample correlation between the antecedent and MLAAR. The dashed line marks the 0.50 criterion.

Appendix B

Table A2 lists, for the outputs and figures reported in this article, the scripts that produce them, the random seeds, the approximate computing time, and the Open Science Framework repository file (https://doi.org/10.17605/OSF.IO/BZUYS) that contains the outputs. Re-running the code on a clean copy of the repository reproduced the published outputs exactly.

Table A2. Reproducibility Summary: Scripts, Seeds, Computing Time, and Repository Files for Each Reported Result.

Results in This Article	Script and Command	Seeds	Approx. Computing Time	Repository File
Tables 9, 10, 12 and 13; Figures 3–5	mlaar_simulation.py --reps 100 --boot 100	s = 20,261,014 + 100,000iN + 10000iρ + 1000if + r (Equation (6))	1.6 CPU-h	results_main.zip

Table A2. *Cont.*

Results in This Article	Script and Command	Seeds	Approx. Computing Time	Repository File
Tables 6 and 8; Figure 6; statistics in Sections 4.1 and 4.2	illustrative_analysis.py	20,261,014 (N = 681, ρ = 0.3, linear; B = 5000)	2 min	illustrative.zip
Table 7; Table A1; Figure A1	hcaio_diagnostics.py	Main-study seeds	1 min	hcaio_diagnostics.zip
Table 11	bootstrap_sensitivity.py --boot 1000	Main-study seeds; resampling stream s + 7,000,000	2 CPU-h	bootstrap_sensitivity.zip
Table 14	estimator_comparison.py --boot 100	Main-study seeds; resampling stream s + 9,000,000	0.5 CPU-h	estimator_comparison.zip
Figures 1 and 2	figures_static.py	–	<1 min	figures.zip
Figures 3–6 (drawing)	figures_results.py results_main	–	<1 min	figures.zip

Appendix C

Table A3 summarises representative studies reviewed in Section 2.5. Table A4 links each hypothesis in Section 2 to its theoretical basis, mechanism, key sources, and population value.

Table A3. Representative Related Studies and Their Evaluation Scope.

Study	Focus	Methods	Known Population Parameters	Structural Accuracy	Predictive Accuracy	Importance Ranking
Liébana-Cabanillas et al. [72]	Mobile-commerce acceptance	SEM + ANN	No	No	Yes	ANN sensitivity
Leong et al. [73]	Mobile-wallet resistance	SEM + ANN	No	No	Yes	ANN sensitivity
Shmueli et al. [75]	Predictive assessment guidelines	PLS-SEM, PLSpredict	No	No	Yes	No
Reinartz et al. [76]	Estimator efficacy	CB-SEM vs. PLS (simulation)	Yes	Yes	No	No
Goodhue et al. [77]	Small samples, non-normal data	PLS, CB-SEM, regression (simulation)	Yes	Yes	No	No
Dijkstra and Henseler [21]	Consistency of PLS	PLS vs. PLSc (simulation)	Yes	Yes	No	No
Sarstedt et al. [78]	Source of PLS and CB-SEM bias	PLS, CB-SEM (simulation)	Yes	Yes	No	No
Yuan et al. [23]	Mediation precision	SEM vs. weighted composites	Yes	Yes (mediation)	No	No
Kumar et al. [80]	Shapley importance	Shapley analysis, constructed examples	Yes	No	No	Yes
Aas et al. [27]	Shapley values with dependent features	Conditional Shapley estimation (simulation)	Yes	No	No	Yes
Present study	Human–AI SDSS (MLAAR)	PLS-PM, PLSc, OLS, GBM, ANN, SHAP (simulation)	Yes	Yes (incl. mediation)	Yes	Yes (vs. population ranking)

Table A4. Theoretical Substantiation of the Hypotheses and Their Population Values.

Hypothesis	Path (Sign)	Theoretical Basis	Mechanism	Key Sources	Population Value
H1	AIAA → MLAAR (+)	Levels of automation; status quo bias	Loss of control and fewer opportunities for intervention	[29,37,43,44]	0.24
H2	AIDO → MLAAR (+)	Algorithmic opacity; explanation research	Inability to verify and justify recommendations; accountability gap	[45–48,50]	0.26
H3	PDAT → MLAAR (+)	User-resistance theory (threat to power)	Perceived loss of discretion, status, and authority	[29,30,51,53]	0.37
H4	HAIT → MLAAR (−)	Trust in automation; organisational trust	Willingness to rely on AI under uncertainty	[8,42,54,55]	−0.18
H5	AIGO → MLAAR (−)	Responsible AI governance	Clear accountability, monitoring, and override reduce uncertainty	[44,57,58,60]	−0.09
H6	HCAIO → MLAAR (−)	Human-centred AI; Industry 5.0	Augmentation framing reduces authority threat	[2,3,36,62]	−0.21
H7	MLAAR → SDMSE (−)	Hybrid intelligence; strategic decision effectiveness	Underuse of AI advice; weaker human–AI complementarity	[6,9,32,64]	a = −0.40
H8	SDMSE → IR (+)	Capability-based organisational resilience	Better anticipation, coping, and adaptation	[65,66]	b = 0.45
H9	MLAAR → SDMSE → IR (−)	Partial-mediation logic	Resistance lowers resilience partly through decision effectiveness	[67]	ab = −0.18 ($c' = -0.10$)

References

1. Csaszar, F.A.; Ketkar, H.; Kim, H. Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. *Strategy Sci.* **2024**, *9*, 322–345. [CrossRef]
2. Breque, M.; De Nul, L.; Petridis, A. *Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry*; Publications Office of the European Union: Luxembourg, 2021.
3. Shneiderman, B. Human-centred artificial intelligence: Reliable, safe and trustworthy. *Int. J. Hum.-Comput. Interact.* **2020**, *36*, 495–504. [CrossRef]
4. Shrestha, Y.R.; Ben-Menahem, S.M.; von Krogh, G. Organizational decision-making structures in the age of artificial intelligence. *Calif. Manag. Rev.* **2019**, *61*, 66–83. [CrossRef]
5. Seeber, I.; Bittner, E.; Briggs, R.O.; de Vreede, T.; de Vreede, G.-J.; Elkins, A.; Maier, R.; Merz, A.B.; Oeste-Reiß, S.; Randrup, N.; et al. Machines as teammates: A research agenda on AI in team collaboration. *Inf. Manag.* **2020**, *57*, 103174. [CrossRef]
6. Dietvorst, B.J.; Simmons, J.P.; Massey, C. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *J. Exp. Psychol. Gen.* **2015**, *144*, 114–126. [CrossRef] [PubMed]
7. Maddikunta, P.K.R.; Pham, Q.-V.; Prabadevi, B.; Deepa, N.; Dev, K.; Gadekallu, T.R.; Ruby, R.; Liyanage, M. Industry 5.0: A survey on enabling technologies and potential applications. *J. Ind. Inf. Integr.* **2022**, *26*, 100257. [CrossRef]
8. Glikson, E.; Woolley, A.W. Human trust in artificial intelligence: Review of empirical research. *Acad. Manag. Ann.* **2020**, *14*, 627–660. [CrossRef]
9. Dellermann, D.; Ebel, P.; Söllner, M.; Leimeister, J.M. Hybrid intelligence. *Bus. Inf. Syst. Eng.* **2019**, *61*, 637–643. [CrossRef]

10. Baird, A.; Maruping, L.M. The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS Q.* **2021**, *45*, 315–341. [CrossRef]
11. Jarrahi, M.H. Artificial intelligence and the future of work: Human–AI symbiosis in organisational decision making. *Bus. Horiz.* **2018**, *61*, 577–586. [CrossRef]
12. Alomair, G.; Salem, M.A. Governing risks arising from institutional AI-driven decision support practices: A case of women leaders in tertiary education. *Algorithms* **2026**, *19*, 832. [CrossRef]
13. Salem, M.A. How AI empowers the double-helix approach: Examining a decision support framework for personalised learning trajectories in gifted education. *Educ. Sci.* **2026**, *16*, 1441. [CrossRef]
14. Salem, M.A. An explainable educational data mining framework for misconception detection and differentiated instruction. *Algorithms* **2026**, *19*, 719. [CrossRef]
15. Salem, M.A.; Khalil, Z.A. Integrating semantic NLP and PLS-SEM for AI-enabled strategic decision support: An explainable framework for assessing organisational AI illiteracy. *Information* **2026**, *17*, 815. [CrossRef]
16. Hair, J.F.; Risher, J.J.; Sarstedt, M.; Ringle, C.M. When to use and how to report the results of PLS-SEM. *Eur. Bus. Rev.* **2019**, *31*, 2–24. [CrossRef]
17. Shmueli, G. To explain or to predict? *Stat. Sci.* **2010**, *25*, 289–310. [CrossRef]
18. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
19. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30*; Curran Associates: Red Hook, NY, USA, 2017; pp. 4765–4774.
20. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef] [PubMed]
21. Dijkstra, T.K.; Henseler, J. Consistent partial least squares path modeling. *MIS Q.* **2015**, *39*, 297–316. [CrossRef]
22. Schuberth, F.; Schamberger, T.; Henseler, J. More powerful parameter tests? No, rather biased parameter estimates. Some reflections on path analysis with weighted composites. *Behav. Res. Methods* **2024**, *56*, 4205–4215. [CrossRef] [PubMed]
23. Yuan, K.-H.; Zhang, Z.; Wang, L. Signal-to-noise ratio in estimating and testing the mediation effect: Structural equation modeling versus path analysis with weighted composites. *Psychometrika* **2024**, *89*, 974–1006. [CrossRef] [PubMed]
24. Grinsztajn, L.; Oyallon, E.; Varoquaux, G. Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems 35*; Neural Information Processing Systems Foundation: New Orleans, LA, USA, 2022; pp. 507–520. [CrossRef]
25. Christodoulou, E.; Ma, J.; Collins, G.S.; Steyerberg, E.W.; Verbakel, J.Y.; Van Calster, B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* **2019**, *110*, 12–22. [CrossRef] [PubMed]
26. Hu, X.; Zhu, M.; Feng, Z.; Stanković, L. Manifold-based Shapley explanations for high dimensional correlated features. *Neural Netw.* **2024**, *180*, 106634. [CrossRef] [PubMed]
27. Aas, K.; Jullum, M.; Løland, A. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artif. Intell.* **2021**, *298*, 103502. [CrossRef]
28. Morris, T.P.; White, I.R.; Crowther, M.J. Using simulation studies to evaluate statistical methods. *Stat. Med.* **2019**, *38*, 2074–2102. [CrossRef] [PubMed]
29. Kim, H.-W.; Kankanhalli, A. Investigating user resistance to information systems implementation: A status quo perspective. *MIS Q.* **2009**, *33*, 567–582. [CrossRef]
30. Lapointe, L.; Rivard, S. A multilevel model of resistance to information technology implementation. *MIS Q.* **2005**, *29*, 461–491. [CrossRef]
31. Castelo, N.; Bos, M.W.; Lehmann, D.R. Task-dependent algorithm aversion. *J. Mark. Res.* **2019**, *56*, 809–825. [CrossRef]
32. Burton, J.W.; Stein, M.-K.; Jensen, T.B. A systematic review of algorithm aversion in augmented decision making. *J. Behav. Decis. Mak.* **2020**, *33*, 220–239. [CrossRef]
33. Mahmud, H.; Islam, A.K.M.N.; Ahmed, S.I.; Smolander, K. What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technol. Forecast. Soc.* **2022**, *175*, 121390. [CrossRef]
34. Hughes, L.; Dwivedi, Y.K.; Malik, T.; Shawosh, M.; Albashrawi, M.A.; Jeon, I.; Dutot, V.; Appanderanda, M.; Crick, T.; De, R.; et al. AI agents and agentic systems: A multi-expert analysis. *J. Comput. Inf. Syst.* **2025**, *65*, 489–517. [CrossRef]
35. Murray, A.; Rhymer, J.; Sirmon, D.G. Humans and technology: Forms of conjoined agency in organizations. *Acad. Manag. Rev.* **2021**, *46*, 552–571. [CrossRef]
36. Raisch, S.; Krakowski, S. Artificial intelligence and management: The automation–augmentation paradox. *Acad. Manag. Rev.* **2021**, *46*, 192–210. [CrossRef]
37. Langer, M.; Landers, R.N. The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Comput. Hum. Behav.* **2021**, *123*, 106878. [CrossRef]

38. Twenge, J.M.; Campbell, S.M.; Hoffman, B.J.; Lance, C.E. Generational differences in work values: Leisure and extrinsic values increasing, social and intrinsic values decreasing. *J. Manag.* **2010**, *36*, 1117–1142. [CrossRef]
39. Logg, J.M.; Minson, J.A.; Moore, D.A. Algorithm appreciation: People prefer algorithmic to human judgment. *Organ. Behav. Hum. Decis. Process.* **2019**, *151*, 90–103. [CrossRef]
40. Puranam, P. Human–AI collaborative decision-making as an organization design problem. *J. Organ. Des.* **2021**, *10*, 75–80. [CrossRef]
41. Amershi, S.; Weld, D.; Vorvoreanu, M.; Fourney, A.; Nushi, B.; Collisson, P.; Suh, J.; Iqbal, S.; Bennett, P.N.; Inkpen, K.; et al. Guidelines for human–AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–13. [CrossRef]
42. Lee, J.D.; See, K.A. Trust in automation: Designing for appropriate reliance. *Hum. Factors* **2004**, *46*, 50–80. [CrossRef] [PubMed]
43. Parasuraman, R.; Sheridan, T.B.; Wickens, C.D. A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cybern. A Syst. Hum.* **2000**, *30*, 286–297. [CrossRef] [PubMed]
44. Dietvorst, B.J.; Simmons, J.P.; Massey, C. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Manag. Sci.* **2018**, *64*, 1155–1170. [CrossRef]
45. Burrell, J. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data Soc.* **2016**, *3*, 1–12. [CrossRef]
46. Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.* **2019**, *267*, 1–38. [CrossRef]
47. Rai, A. Explainable AI: From black box to glass box. *J. Acad. Mark. Sci.* **2020**, *48*, 137–141. [CrossRef]
48. Shin, D. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *Int. J. Hum.-Comput. Stud.* **2021**, *146*, 102551. [CrossRef]
49. Alufaisan, Y.; Marusich, L.R.; Bakdash, J.Z.; Zhou, Y.; Kantarcioglu, M. Does explainable artificial intelligence improve human decision-making? *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 6618–6626. [CrossRef]
50. Grimmelikhuijsen, S. Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Adm. Rev.* **2023**, *83*, 241–262. [CrossRef]
51. Faraj, S.; Pachidi, S.; Sayegh, K. Working and organizing in the age of the learning algorithm. *Inf. Organ.* **2018**, *28*, 62–70. [CrossRef]
52. Dwivedi, Y.K.; Hughes, L.; Ismagilova, E.; Aarts, G.; Coombs, C.; Crick, T.; Duan, Y.; Dwivedi, R.; Edwards, J.; Eirug, A.; et al. Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int. J. Inf. Manag.* **2021**, *57*, 101994. [CrossRef]
53. Kellogg, K.C.; Valentine, M.A.; Christin, A. Algorithms at work: The new contested terrain of control. *Acad. Manag. Ann.* **2020**, *14*, 366–410. [CrossRef]
54. Mayer, R.C.; Davis, J.H.; Schoorman, F.D. An integrative model of organizational trust. *Acad. Manag. Rev.* **1995**, *20*, 709–734. [CrossRef]
55. Wen, Y.; Wang, J.; Chen, X. Trust and AI weight: Human-AI collaboration in organizational management decision-making. *Front. Organ. Psychol.* **2025**, *3*, 1419403. [CrossRef]
56. Wu, Y.; Zhang, J. A Hybrid Human–AI Consensus Model in Information Granularity-Based Large-Scale Group Decision-Making. *Mathematics* **2026**, *14*, 3506. [CrossRef]
57. Laux, J. Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of AI governance under the European Union AI Act. *AI Soc.* **2024**, *39*, 2853–2866. [CrossRef] [PubMed]
58. Mäntymäki, M.; Minkkinen, M.; Birkstedt, T.; Viljanen, M. Defining organizational AI governance. *AI Ethics* **2022**, *2*, 603–609. [CrossRef]
59. Engin, Z. Human-AI governance (HAIG): A trust-utility approach. *J. Responsible Technol.* **2026**, *26*, 100167. [CrossRef]
60. Papagiannidis, E.; Mikalef, P.; Conboy, K. Responsible artificial intelligence governance: A review and research framework. *J. Strateg. Inf. Syst.* **2025**, *34*, 101885. [CrossRef]
61. Schmitz, C.; Rystrøm, J.; Batzner, J. Oversight structures for agentic AI in public-sector organizations. *arXiv* **2025**, arXiv:2506.04836. [CrossRef]
62. Xu, X.; Lu, Y.; Vogel-Heuser, B.; Wang, L. Industry 4.0 and Industry 5.0—Inception, conception and perception. *J. Manuf. Syst.* **2021**, *61*, 530–535. [CrossRef]
63. Golgeci, I.; Ritala, P.; Arslan, A.; McKenna, B.; Ali, I. Confronting and alleviating AI resistance in the workplace: An integrative review and a process framework. *Hum. Resour. Manag. Rev.* **2025**, *35*, 101075. [CrossRef]
64. Elbanna, S.; Child, J. Influences on strategic decision effectiveness: Development and test of an integrative model. *Strateg. Manag. J.* **2007**, *28*, 431–453. [CrossRef]
65. Duchek, S. Organizational resilience: A capability-based conceptualization. *Bus. Res.* **2020**, *13*, 215–246. [CrossRef]
66. Linnenluecke, M.K. Resilience in business and management research: A review of influential publications and a research agenda. *Int. J. Manag. Rev.* **2017**, *19*, 4–30. [CrossRef]

67. Zhao, X.; Lynch, J.G., Jr.; Chen, Q. Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *J. Consum. Res.* **2010**, *37*, 197–206. [CrossRef]
68. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.
69. Cheng, Q.; Dai, Y.; Liu, X.; Peng, S. The trust crisis in artificial intelligence: AI hallucinations and human-AI collaboration. *Technol. Soc.* **2026**, *86*, 103286. [CrossRef]
70. Jin, S.; Roh, S. Effects of AI explanations on human-AI collaboration: An experimental study on decision performance and reliance. *Seoul J. Bus.* **2025**, *31*, 67–99. [CrossRef]
71. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009. [CrossRef]
72. Liébana-Cabanillas, F.; Marinković, V.; Kalinić, Z. A SEM-neural network approach for predicting antecedents of m-commerce acceptance. *Int. J. Inf. Manag.* **2017**, *37*, 14–24. [CrossRef]
73. Leong, L.-Y.; Hew, T.-S.; Ooi, K.-B.; Wei, J. Predicting mobile wallet resistance: A two-staged structural equation modeling-artificial neural network approach. *Int. J. Inf. Manag.* **2020**, *51*, 102047. [CrossRef]
74. Hofman, J.M.; Watts, D.J.; Athey, S.; Garip, F.; Griffiths, T.L.; Kleinberg, J.; Margetts, H.; Mullainathan, S.; Salganik, M.J.; Vazire, S.; et al. Integrating explanation and prediction in computational social science. *Nature* **2021**, *595*, 181–188. [CrossRef] [PubMed]
75. Shmueli, G.; Sarstedt, M.; Hair, J.F.; Cheah, J.-H.; Ting, H.; Vaithilingam, S.; Ringle, C.M. Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *Eur. J. Mark.* **2019**, *53*, 2322–2347. [CrossRef]
76. Reinartz, W.; Haenlein, M.; Henseler, J. An empirical comparison of the efficacy of covariance-based and variance-based SEM. *Int. J. Res. Mark.* **2009**, *26*, 332–344. [CrossRef]
77. Goodhue, D.L.; Lewis, W.; Thompson, R. Does PLS have advantages for small sample size or non-normal data? *MIS Q.* **2012**, *36*, 981–1001. [CrossRef]
78. Sarstedt, M.; Hair, J.F.; Ringle, C.M.; Thiele, K.O.; Gudergan, S.P. Estimation issues with PLS and CBSEM: Where the bias lies! *J. Bus. Res.* **2016**, *69*, 3998–4010. [CrossRef]
79. Rönkkö, M.; McIntosh, C.N.; Antonakis, J.; Edwards, J.R. Partial least squares path modeling: Time for some serious second thoughts. *J. Oper. Manag.* **2016**, *47–48*, 9–27. [CrossRef]
80. Kumar, I.E.; Venkatasubramanian, S.; Scheidegger, C.; Friedler, S. Problems with Shapley-value-based explanations as feature importance measures. In Proceedings of the 37th International Conference on Machine Learning, Virtual, 13–18 July 2020; PMLR 119, pp. 5491–5500.
81. Henseler, J.; Ringle, C.M.; Sarstedt, M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J. Acad. Mark. Sci.* **2015**, *43*, 115–135. [CrossRef]
82. Fornell, C.; Larcker, D.F. Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* **1981**, *18*, 39–50. [CrossRef]
83. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman & Hall: New York, NY, USA, 1993.
84. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.