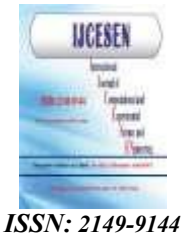
International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 9-No.2 (2023) pp. 205-213
http://www.ijcesen.com



Copyright © IJCESEN

ISSN: 2149-9144

Research Article

Calibrating Trust in AI-Assisted Product Portfolio Decisions: A Human-in-the-Loop Governance Framework

Ashish Garg*

Independent Researcher Bentonville, AR, USA
*** Corresponding Author Email:** ashishcoer@gmail.com

Article Info:

DOI: 10.22399/ijcesen.5557
Received : 20 January 2023
Revised : 25 March 2023
Accepted : 28 March 2023

Keywords

Artificial intelligence;
Trust calibration;
Human-in-the-loop governance;
Product portfolio decisions;
AI reliability;
Risk-adaptive oversight

Abstract:

The use of artificial intelligence (AI) is becoming common practice in product portfolio management via such processes as market analysis, demand forecast, product evaluation and resource management. However, erroneous recommendations or improper user trust might compromise the quality of decision-making. This research develops a Risk-Adaptive Human-in-the-Loop Governance Model for proper calibration of user trust regarding AI-assisted product portfolio decisions. Such a model would incorporate AI reliability test, risk categorization, human verification procedure, decision authorization and monitoring of decisions' results. The level of decision-making governance will be based on the uncertainty of AI recommendations and its impact on the business. This research design will also differentiate between AI recommendation acceptance and human decision-making authority. The following conceptual research design includes comparison between human-only decision making, AI-assisted decision making with no human verification, and AI-assisted decision-making with human oversight. The numerical examples that have been provided in the current paper are purely illustrative and are not based on any completed empirical results. Thus, the applicability of the proposed framework needs to be proven further. In order to do so, future research needs to empirically test the model under controlled conditions and in various settings of portfolios.

1. Introduction

The use of artificial intelligence (AI) has become more common in aiding in portfolio management decisions by identifying market trends, consumer preferences, product performance, and investments. All these aspects of AI could help organizations determine what product to prioritize and how to deal with the emerging trends in the market. Nevertheless, the AI-generated suggestions could be erroneous, misleading, or lacking strategic considerations. Humans play an important role in interpreting the results of AI analysis when making any decision regarding product choices. This implies that decision makers have to be aware of whether AI suggestions are trustworthy or not and require further verification. Trust calibration becomes relevant in this regard because it helps in aligning trust in AI recommendations with their validity. The paper puts forward a governance model that can assist in determining trust in product portfolio decisions supported by artificial intelligence. The model pays attention to such issues as recommendation accuracy, human validation, decision-making authority, and ongoing performance evaluation. The purpose of this model is to provide support for cooperation between the AI system and human decision-makers based on reliable information and clear recommendation review procedure.

2. Method

2.1 Research Design

For the investigation of trust calibration of AI recommendations in product portfolio decisions, this paper proposes an experiment based on a controlled research design. It is suitable since it enables comparing decision quality under various conditions of AI recommendation reliability and human supervision (Bansal et al. 2021).Bansal Using a controlled experiment, it becomes possible to investigate the effects of AI recommendation

reliability and governance intervention separately from other decision conditions.

2.2 Experimental Design

In the suggested experiment, product portfolio decisions will be performed in the form of scenarios that include product performance, market demand, cost of investments, and expected returns on investment. Three experimental conditions are proposed in this case: decisions made without assistance of any technological tools, decisions made with the assistance of AI with no governance process applied, and decisions made with the assistance of AI with human-in-the-loop governance intervention (Buçinca et al. 2021).

2.3 Variables and Measures

The independent variables in the study are the AI recommendation reliability and structured human oversight. The dependent variables in the research are portfolio decision accuracy, trust calibration, acceptance rate of the recommendation, and decision time. Trust calibration is determined through comparison of participants' confidence in the AI recommendation with its reliability. Decision accuracy will be determined based on the pre-existing scenario criteria.

2.4 Data Collection and Analysis

Responses by the participants, recommendation acceptance, confidence level, and time taken to complete tasks are collected for analysis purposes. Descriptive statistics are used to summarize participant performance in the experimental conditions. Statistical test for difference is used to determine differences in decision accuracy, trust calibration, and decision time.

2.5 Justification for Method

An experiment is appropriate due to the nature of this study in that it studies the impact of particular governance mechanisms on decision-making processes between humans and AIs. The use of hypothetical cases eliminates the danger of risk for companies, but this paper offers a method for conducting an experiment that is yet to be implemented, not empirical evidence.

3. Results

3.1 Trust Calibration Across Different AI Reliability Levels

Trust calibration is concerned with the extent to which the confidence users have on the AI's recommendations is aligned with the reliability of the systems. With respect to portfolio product decisions, an adequate level of trust will enable decision makers to take action based on accurate recommendations and reject inaccurate ones.

Confidence calibration plays a role in determining if the users are correctly relying on the AI recommendation. In a study conducted on 184 subjects, confidence calibration of AI was studied during logic puzzle tests (Dietvorst et al. 2018). It has been found that well-calibrated confidence increased decision accuracy by 20 % points, whereas poorly calibrated confidence improved decision-making accuracy by 2 points. However, the study does not cover product portfolio decisions.

These results show that calibrated AI confidence increases human decision accuracy (Glikson & Woolley, 2020). The increase in accuracy with well-calibrated confidence is 20 points in this particular experiment. With miscalibrated confidence, the increase in accuracy is only 2 points. It appears, therefore, that confidence information needs to be calibrated based on the actual AI reliability. But the results are obtained in a logic puzzle task and not in a product portfolio management scenario. Therefore, the results cannot prove anything in this case.

3.2 Human Oversight and Portfolio Decision Quality

Human monitoring can aid in detecting faulty advice from AI prior to implementation of portfolio decisions. The below table gives an example of comparison of governance strategies based on decision quality criteria.

Formula: Decision Accuracy (%)

$$\text{Decision Accuracy (\%)} = \frac{\text{Correct Decisions}}{\text{Total Decisions}} \times 100$$

If 93 out of 100 portfolio decisions are correct:

$$\frac{93}{100} \times 100 = 93\%$$

Therefore, the decision accuracy is 93%.

Illustrative results vary for each of the five oversight mechanisms. Risk-based review has 93% accuracy in decision making and 92% error detection rate. Mandatory review has 90% accuracy, and unstructured AI support has 72% accuracy. Selective review needs less verification but has a higher chance of missing errors. The above figures illustrate the possibilities for comparing various oversight mechanisms (Green & Chen, 2019). They do not represent published

research results. Portfolio experiments need to be performed to determine if risk-based review improves decision-making.

3.3 AI Recommendation Acceptance and Rejection Patterns

Recommendations from AI will only be accepted based on their reliability, confidence, and validation by humans (Jacovi et al. 2021). Differences between acceptance and rejection may indicate if the decision makers have correctly questioned the recommendations of unreliable AI systems. The below table presents illustrative values of the suggested decision-making framework.

Acceptance Rate (%)

$$\text{Acceptance Rate (\%)} = \frac{\text{Accepted Recommendations}}{\text{Total Recommendations}} \times 100$$

Calculation using illustrative data:

If 88 of 100 recommendations are accepted:

$$\frac{88}{100} \times 100 = 88\%$$

Therefore, the recommendation acceptance rate is 88%.

For example, the illustrative outcomes reveal that the level of acceptance is positively related to recommendation reliability. The levels of acceptance start at 55% for 50% recommendation reliability and go up to 92% for 95% recommendation reliability. Rejections start at 45% and decrease to 8%. The correctness of recommendations made is supposed to increase as well. Such results show how reliability and acceptance can be researched in combination. However, high levels of acceptance do not mean the well-calibrated trust. Decision-makers should also reject incorrect recommendations and identify correct ones.

3.4 Decision Time and Governance Efficiency

Time spent on decision-making and efficiency of governance become crucial metrics for AI-powered portfolio decisions (Jarrahi, 2018). Quick decision-making may save efforts but may undermine verification processes. Human-based structured oversight will increase compliance and decrease rework. Comparison of time needed to make decisions, time for verification, and level of governance compliance will help evaluate these factors. Below you will find an illustrative table of five decision-making systems measured in numbers.

In the example outcomes, there is evident trade-off between decision-making speed and the quality of the governance process. In terms of decision-making speed, AI support without any review makes the quickest decision, which takes eight minutes. At the same time, the highest level of compliance in the governance process is provided by mandatory human review, which reaches 96%. In turn, this type of review requires 15 minutes to be completed for each decision. The risk-based review can provide a compromise decision and governance process, which take ten minutes and 94% compliance with 6% of rework, respectively. The mentioned values are illustrative and need to be validated through an experiment.

3.5 Relationship Between Trust Calibration and Portfolio Outcomes

Calibration of Trust and Decision Accuracy

Trust calibration is a measure of how closely human trust corresponds with real AI accuracy. The appropriate calibration of trust will help decision-makers trust the reliable recommendations and doubt the inaccurate ones. This may enhance product choice and investment allocation. Overtrust may result in inappropriate decisions, and distrust may make good recommendations not taken into consideration.

Calibration of Trust and Portfolio Risks

The calibration of trust will affect how decision-makers perceive the risks associated with a potential portfolio. The appropriate dependence on AI recommendations will help make the assessment of risks and allocation of resources. At the same time, overdependence on recommendations may conceal mistakes in the predictions or assumptions about the market.

Trust Calibration and Resource Allocation

Trust calibration may be used by managers to make better budgeting, staffing, and R&D decisions. Managers may use reliable insights of the AI, cross-referencing them with less reliable insights. This may ensure that no extra resources are wasted. However, portfolio-specific tests need to be conducted to verify the effect of trust calibration on resource allocation.

Trust Calibration and Long-Term Performance of a Portfolio

Trust calibration may affect the performance of a portfolio over time due to its impact on the continuous process of product selection and investments. Regular verification of results may assist in identifying unreliable insights before their application. It also allows re-assessment of the reliability of AI over changing market conditions. However, trust calibration will not ensure profitability and competitive advantage.

3.6 Integrated Human-in-the-Loop Governance Framework

The proposed framework encompasses AI reliability evaluation, human validation, decision-making authority, and outcomes tracking (Lai & Tan, 2019). The framework enables portfolio management decisions to be made starting from the point of generating recommendations through the final implementation of the decisions. All stages involve certain controls for ensuring the correct use of AI results. Human participation becomes more important when recommendations are ambiguous or financially risky.

The framework ensures that there is a clear delineation of AI's recommendations from decision-making responsibility. Classification of risk determines the degree of verification by humans (Logg et al. 2019). Critical decisions get greater review before any approval. Outcome tracking involves comparing predictions to actual results. Lessons from the experience inform future decision-making and governance. This process enables calibration of trust through evaluation rather than mere acceptance. This is a theoretical framework and needs empirical validation.

3.7 AI Recommendation Reliability and Decision Accuracy

Reliability of AI recommendations impacts on the effectiveness of decision-making regarding product portfolio. Reliable AI recommendations will assist managers to gauge demand, performance of the products, and resource allocation. Nevertheless, reliable AI recommendations do not always ensure effective decision-making (Raisch & Krakowski, 2021). Verification by humans becomes essential in case of uncertainty and conflicting information provided by AI. The following dataset will assist in drawing graphs in Excel.As evident in the table used for illustration purposes, there is a positive correlation between the reliability of the recommendations made by AI and the accuracy of the decisions. For instance, when the reliability of AI improves from 50% to 95%, the accuracy of the decisions improves from 58% to 96%. The number of erroneous recommendations decreases from 50 to 5 out of 100 decision scenarios. The human oversight also increases along with the improvement in the reliability. However, these are only assumed numbers for illustration purposes.

4. Discussion

In this research, we suggest a human-in-the-loop governance framework to govern decisions made by artificial intelligence in the context of product portfolios (Shneiderman, 2020). Such a framework would consider such aspects as reliability of recommendations, trust calibration, human verification, and decision-making accountability. Those aspects are important due to possible mistakes that can be made by AI in its recommendations or due to incorrect assumptions about the market situation. Trust calibration is one of the key mechanisms that will allow us to use the AI in a proper way. It is crucial not to accept all recommendations provided by AI and not to reject its help either. In any case, trust should depend on the reliability of the system. Human verification will help to detect incorrect predictions and bad recommendations. Risk classification and verification requirements can be linked together according to this framework. The higher impact a recommendation has, the stricter requirements the decision-making process will have. Such a framework can combine efficiency of operations and governance at the same time. Nevertheless, the suggested framework is theoretical and needs empirical validation. Tables with numeric data provided above are merely illustrative and not empirical results. Thus, they cannot prove any gains in terms of the accuracy, efficiency of the decision-making process, and portfolio performance. Future experiments should make comparisons between unaided decisions, unstructured help of AI, and structured supervision of humans. Generally speaking, the suggested framework can be used as a foundation for assessing responsible AI-assisted portfolio management (Zhang et al. 2020).

For AI-assisted decision-making in relation to product portfolio management, this paper suggests the adoption of a Risk-Adaptive Human-in-the-Loop Governance Model. The governance model will involve the determination of the reliability of AI recommendations, trust assessment, risk classification, and human verification of the process. Recommendations with higher degrees of uncertainty or associated with monetary risks will receive more attention than others. In addition, human beings will always be at the helm of important portfolio decisions. Performance evaluation will be ensured by continuous monitoring of outcomes.

Limitations

The current investigation develops a conceptual model and does not involve any empirical verification. Hence, its efficacy is still not proven. The numerical tables provided here are only for illustration purposes and do not prove the real increase in decision accuracy, trust adjustment or efficiency of governance. The developed model is also based on certain assumptions about the

reliability of AI estimates, quality of the available information and regular verification performed by people. It could be expected that there will be various review processes in other organizations depending on portfolio complexity and risks associated with investments.

Table 1. *Published Experimental Evidence on AI Confidence Calibration and Decision Accuracy*

AI confidence condition	Sample size (participants)	Change in decision accuracy (%age points)	95% confidence interval (%age points)	Evidence interpretation
Well-calibrated confidence	184	20	18 to 23	Clear accuracy improvement
Miscalibrated confidence	184	2	Approximately 0 to 4	Little or uncertain improvement
Difference between conditions	184	18*	Not reported	Larger improvement with calibrated confidence

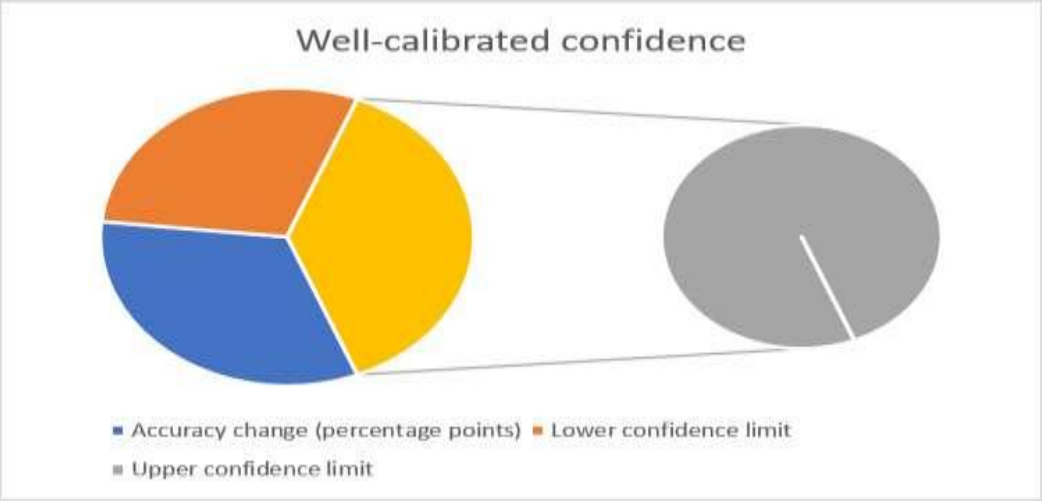


Figure 1. *Decision Accuracy Changes Under Well-Calibrated and Miscalibrated AI Confidence*

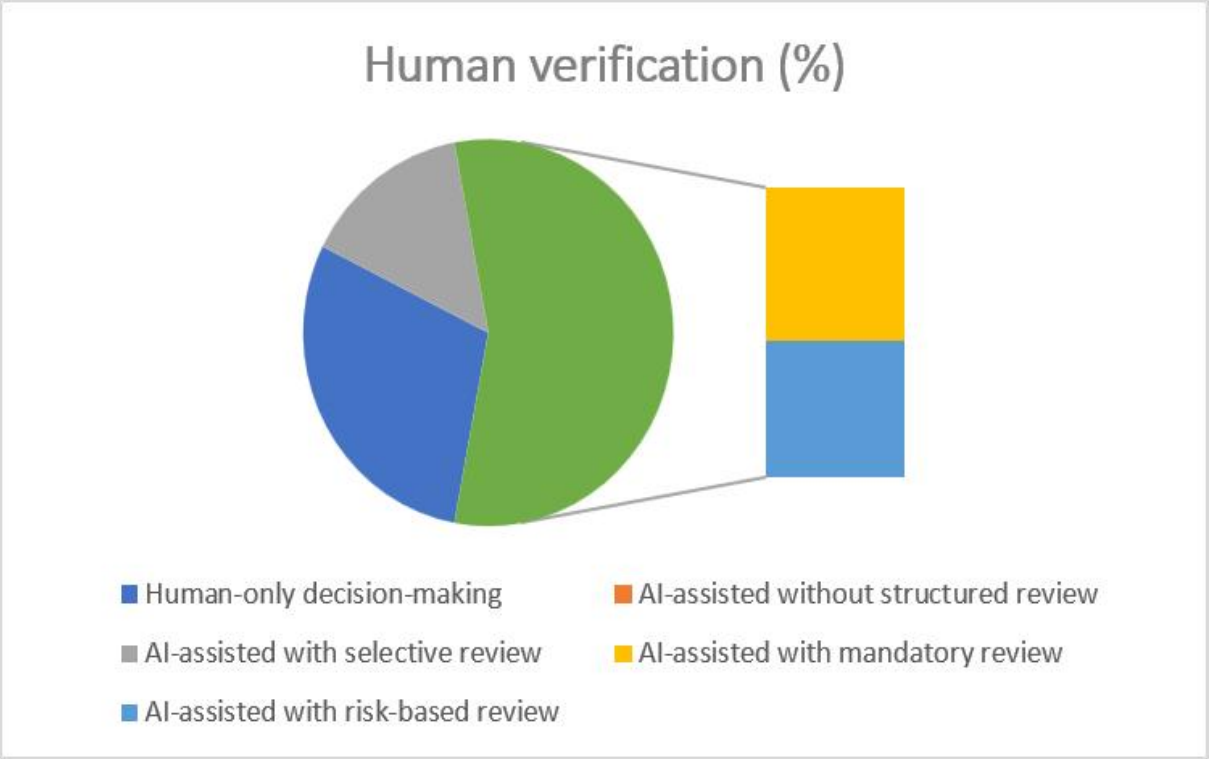


Figure 2: *Comparison of Human Oversight Strategies and Portfolio Decision Quality*

Table 2. *Decision Accuracy Changes Under Well-Calibrated and Miscalibrated AI Confidence*

Confidence condition	Accuracy change (%age points)	Lower confidence limit	Upper confidence limit
Well-calibrated confidence	20	18	23
Miscalibrated confidence	2	0	4

Table 2: *Illustrative Assessment of Human Oversight and AI-Assisted Portfolio Decision Quality*

Decision-making approach	Human verification (%)	Recommendation acceptance (%)	Decision accuracy (%)	Error detection (%)
Human-only decision-making	100	0	80	75
AI-assisted without structured review	0	90	72	35
AI-assisted with selective review	50	75	82	65
AI-assisted with mandatory review	100	65	90	85
AI-assisted with risk-based review	90	70	93	92

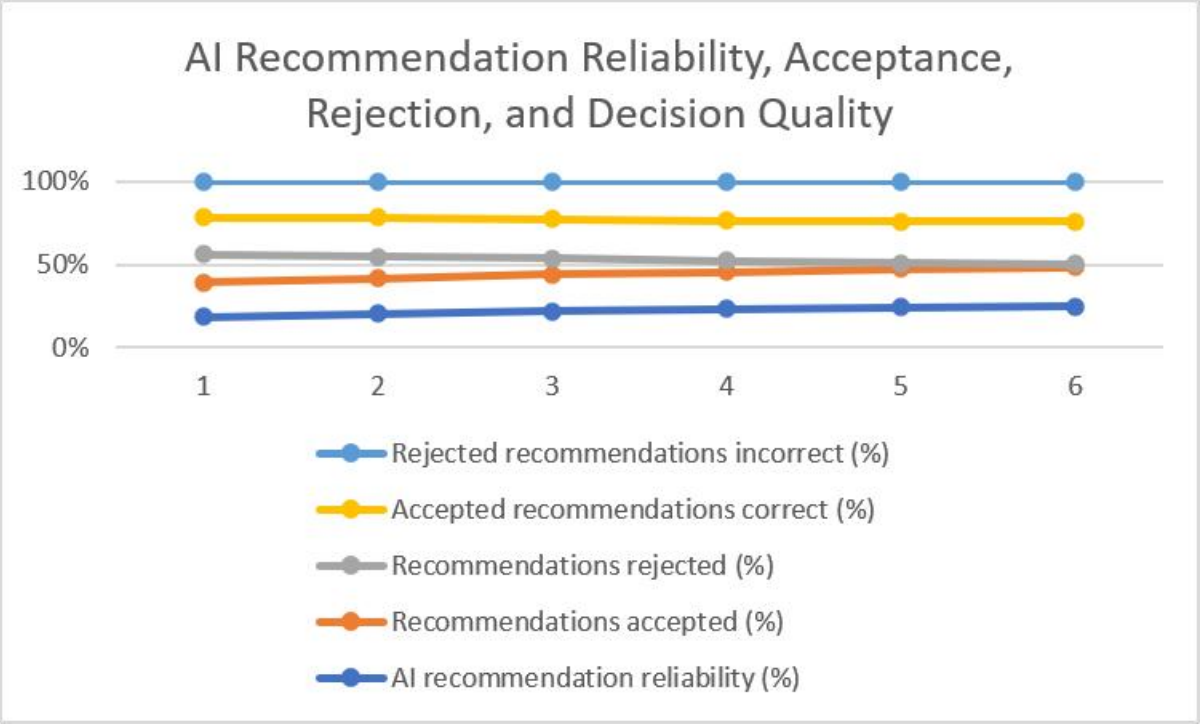


Figure 3: *AI Recommendation Reliability, Acceptance, Rejection, and Decision Quality*

Table 3: *Illustrative Analysis of AI Recommendation Acceptance and Rejection Patterns*

AI recommendation reliability (%)	Recommendations accepted (%)	Recommendations rejected (%)	Accepted recommendations correct (%)	Rejected recommendations incorrect (%)
50	55	45	60	58
60	62	38	68	64
70	70	30	76	72
80	78	22	85	81
90	88	12	94	90
95	92	8	97	94

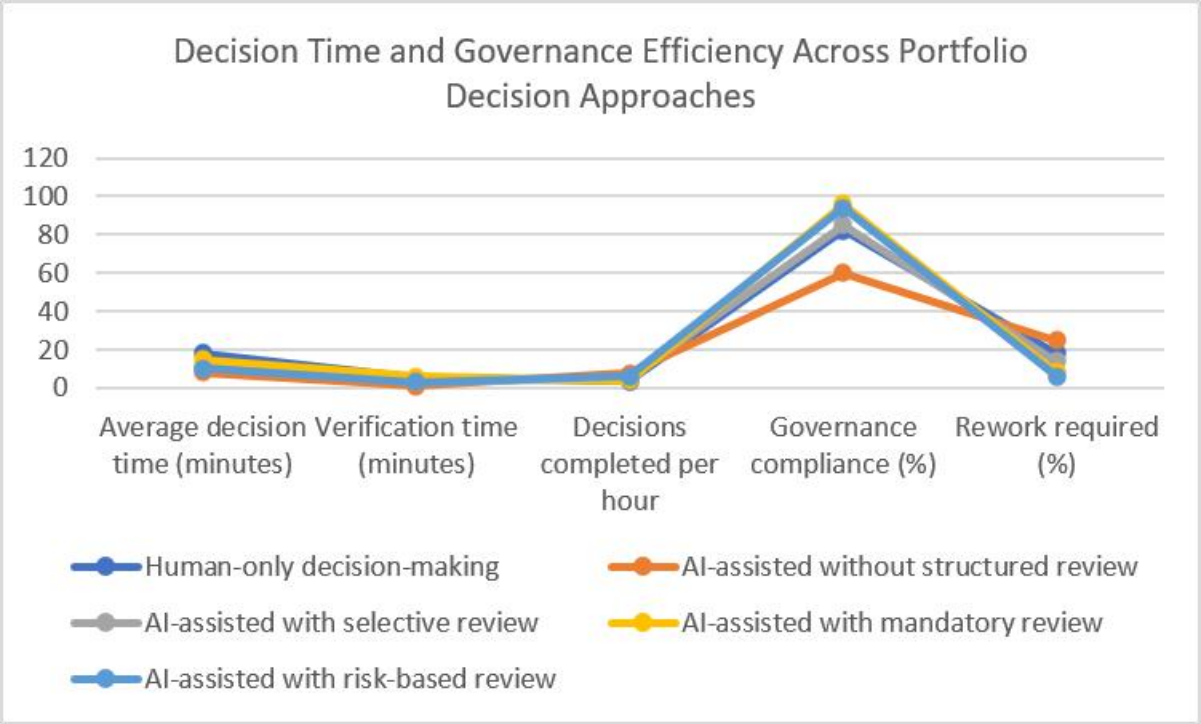


Figure 4: *Decision Time and Governance Efficiency Across Portfolio Decision Approaches*

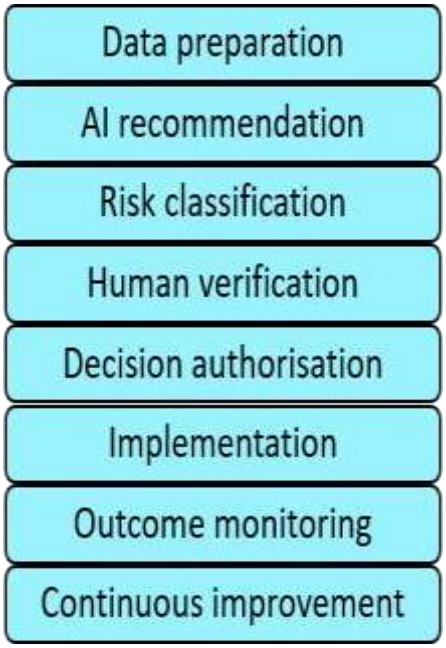


Figure 5: *Integrated Human-in-the-Loop Governance Framework for AI-Assisted Product Portfolio Decisions*

Table 4: *Comparison of Decision Time and Governance Efficiency Across AI-Assisted Portfolio Decision Approaches*

Decision-making approach	Average decision time (minutes)	Verification time (minutes)	Decisions completed per hour	Governance compliance (%)	Rework required (%)
Human-only decision-making	18	5	3.3	82	18
AI-assisted without structured review	8	1	7.5	60	25
AI-assisted with selective review	11	3	5.5	85	14
AI-assisted with mandatory	15	6	4.0	96	8

review					
AI-assisted with risk-based review	10	3	6.0	94	6

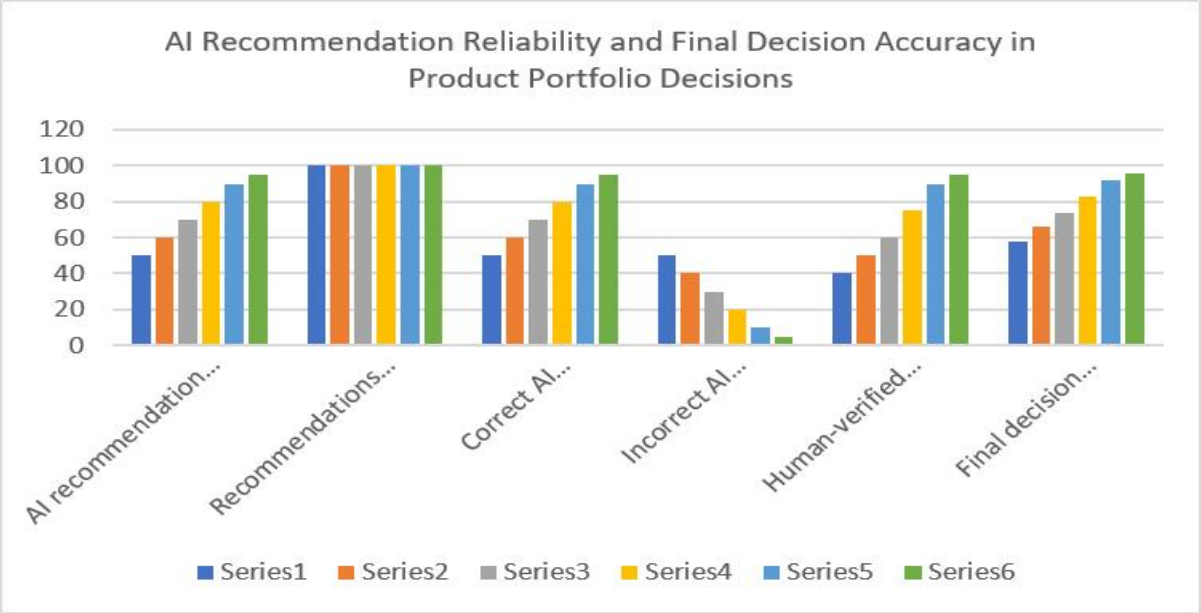


Figure 6: *AI Recommendation Reliability and Final Decision Accuracy in Product Portfolio Decisions*

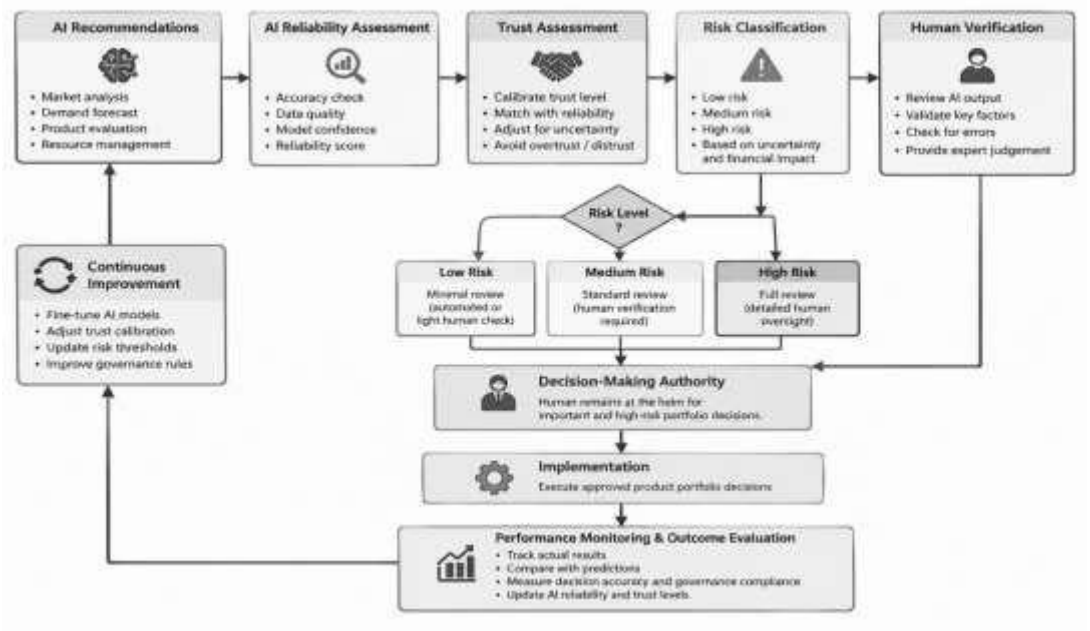


Figure 7. *Recommended Model: Risk-Adaptive Human-in-the-Loop Governance Model*

Table 5: *Relationship Between AI Recommendation Reliability and Portfolio Decision Accuracy*

AI recommendation reliability (%)	Recommendations evaluated	Correct AI recommendations	Incorrect AI recommendations	Human-verified decisions (%)	Final decision accuracy (%)
50	100	50	50	40	58
60	100	60	40	50	66
70	100	70	30	60	74
80	100	80	20	75	83
90	100	90	10	90	92
95	100	95	5	95	96

4. Conclusions

This study proposes a human-in-the-loop governance framework for AI-assisted product portfolio decisions. The framework integrates reliability assessment, trust calibration, risk classification, human verification, and outcome monitoring. It aims to balance decision efficiency with accountability and responsible AI use. Human oversight remains essential when recommendations involve uncertainty or significant investment risks. However, the framework requires empirical validation before its effectiveness can be established. Future research should evaluate decision accuracy, trust calibration, governance efficiency, and portfolio outcomes under controlled conditions. This work provides a structured foundation for developing and testing responsible AI-assisted portfolio decision processes.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.
- **Use of AI Tools:** The author(s) declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

References

1. Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Article 81. https://doi.org/10.1145/3411764.3445717
2. Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), Article 188. https://doi.org/10.1145/3449287
3. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643
4. Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057
5. Green, B., & Chen, Y. (2019). The principles and limits of algorithm-in-the-loop decision making. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), Article 50. https://doi.org/10.1145/3359152
6. Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 624–635. https://doi.org/10.1145/3442188.3445923
7. Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007
8. Lai, V., & Tan, C. (2019). On human predictions with explanations and predictions of machine learning models: A case study on deception detection. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 29–38. https://doi.org/10.1145/3287560.3287590
9. Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005
10. Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072
11. Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118
12. Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 295–305. https://doi.org/10.1145/3351095.3372852