Idea: Future Research
Published by Ideas for Future Research and Technology
ISSN 2987-1565 (Online) ISSN 2987-2715 (Print), Volume 04, Issue 02, July (2026)

Enhancing Decision Accuracy Through Human-in-the-Loop Decision Support Systems: A Human-AI Collaborative Framework

Muhazzim[1], Tahmaseb[2], Atiyyatullah[3]
[1,2] Department of Mass Communication, University of Karachi, Karachi, Pakistan
[3] Department of Computer Science, COMSATS University Islamabad (CUI), Pakistan

Abstrack

The rapid advancement of Artificial Intelligence (AI) has accelerated the adoption of AI-powered Decision Support Systems (DSS) across healthcare, finance, manufacturing, government, and education. Although AI-based DSS enhance decision-making through efficient data analysis and predictive recommendations, fully automated systems remain susceptible to hallucinations, algorithmic bias, misclassification, and limited contextual understanding, which may reduce decision accuracy in high-risk situations. Consequently, Human-in-the-Loop (HITL) has emerged as an important approach for incorporating human oversight into AI-assisted decision-making. However, empirical evidence regarding the effectiveness of HITL in improving decision accuracy remains limited. Therefore, this study investigates the influence of Human-in-the-Loop Decision Support Systems on decision accuracy. A mixed-methods research design was employed, combining quantitative and qualitative approaches. Participants were users performing analytical decision-making tasks using Decision Support Systems. An experimental framework compared AI-only and collaborative Human-AI decision-making using representative datasets. Data were analyzed through descriptive statistics, reliability and validity testing, regression analysis, and hypothesis testing. The results indicate that Human-in-the-Loop Decision Support Systems significantly improve decision accuracy by integrating human expertise with AI capabilities. The proposed Human-AI collaborative framework enhances transparency, reliability, and user trust while providing practical guidance for developing trustworthy Decision Support Systems that balance AI automation with meaningful human participation.

Keyword:
Human-in-the-Loop;
Decision Support System;
Decision Accuracy;
Human-AI Collaboration;
Artificial Intelligence.

Autor correspondence:
[Muhazzim]
[muhazzim@gmail.com]

Received: May 26, 2026
Revise: June 23, 2026
Accepted: July 15, 2026

Introduction

Artificial Intelligence (AI) has emerged as one of the most influential technologies of the Fourth Industrial Revolution, fundamentally transforming how organizations, governments, and individuals process information and make decisions(Schwab & Davis, 2018). The rapid pace of digital transformation has accelerated the integration of AI with big data, cloud computing, intelligent systems, and machine learning, enabling organizations to analyze vast amounts of structured and unstructured data efficiently. Unlike conventional analytical tools, AI systems are capable of identifying hidden patterns, generating predictions, and supporting complex decision-making processes under uncertain conditions. As a result, AI has evolved from a computational technology into an intelligent partner that assists decision-makers in improving the speed, quality, and effectiveness of organizational decisions.

The rapid growth of big data has further increased the demand for intelligent analytical systems capable of converting massive datasets into meaningful knowledge. Organizations continuously

generate data from digital transactions, Internet of Things (IoT) devices, social media, healthcare records, financial systems, and enterprise applications(Behmann & Wu, 2015). Traditional analytical approaches often struggle to process these increasingly large and complex datasets, whereas machine learning algorithms can automatically learn from historical data, recognize patterns, and produce predictive insights. These capabilities have significantly enhanced evidence-based decision-making by allowing organizations to respond more effectively to rapidly changing environments and increasingly complex problems.

The advancement of AI has also transformed the development of Decision Support Systems (DSS). A Decision Support System is a computer-based system designed to assist decision-makers by integrating data, analytical models, and organizational knowledge into a structured decision-making process(Bonczek et al., 2014). Initially, traditional DSS primarily relied on rule-based systems, statistical methods, and manually developed analytical models. Although these systems improved data accessibility and computational efficiency, they possessed limited adaptability because they required extensive human intervention whenever new situations emerged.

Recent technological developments have led to the emergence of Intelligent Decision Support Systems powered by artificial intelligence. These modern DSS employ machine learning, deep learning, natural language processing, computer vision, and predictive analytics to generate adaptive recommendations and automate increasingly sophisticated decision-making tasks. Rather than functioning solely as information providers, AI-powered DSS actively analyze data, identify patterns, estimate probabilities, and recommend optimal alternatives with remarkable speed and scalability. Consequently, intelligent DSS have become essential tools across various sectors, including healthcare, finance, manufacturing, government, and education. In healthcare, AI supports disease diagnosis, treatment recommendations, and patient risk prediction. In finance, AI assists with fraud detection, credit risk analysis, and investment decisions. Manufacturing industries utilize AI for predictive maintenance, quality control, and supply chain optimization, while governments employ intelligent DSS for policy evaluation and resource allocation. Educational institutions increasingly adopt AI to support personalized learning, academic advising, and student performance prediction.

Despite these remarkable advancements, fully autonomous AI systems continue to face important limitations that restrict their ability to make reliable decisions in complex real-world situations(Hagos & Rawat, 2022). Although AI models have achieved impressive predictive performance, they remain vulnerable to several technical and practical challenges. Large Language Models and other AI systems may generate hallucinations by producing plausible but factually incorrect information. Machine learning models may experience overfitting, reducing their ability to generalize to unseen data. AI systems may also misclassify complex cases, fail to interpret contextual information, or generate biased recommendations due to imbalanced training data. These limitations become particularly problematic in high-stakes domains such as healthcare, finance, and public administration, where inaccurate decisions may produce significant economic, ethical, or societal consequences.

These challenges have encouraged researchers to develop Human-in-the-Loop (HITL) Decision Support Systems that combine AI capabilities with human expertise throughout the decision-making process. Human-in-the-Loop refers to an AI framework in which humans actively participate in activities such as monitoring AI outputs, validating recommendations, providing feedback, modifying system decisions, and approving final outcomes. Rather than replacing human decision-makers, HITL emphasizes collaboration between humans and AI by integrating human oversight, intervention, feedback, and verification into intelligent systems. Human oversight ensures that AI recommendations remain consistent with organizational objectives, ethical principles, and contextual realities. Human intervention allows experts to revise AI outputs when unexpected situations arise, while human feedback enables AI systems to improve through continuous learning. Human verification serves as a final quality-

control mechanism before decisions are implemented, particularly in critical applications where errors may have severe consequences.

The primary objective of integrating Human-in-the-Loop into Decision Support Systems is to improve decision accuracy(Wiethof & Bittner, 2021). Decision accuracy refers to the extent to which a decision correctly identifies the most appropriate alternative based on available evidence. High decision accuracy reflects correctness, precision, reliability, consistency, and overall decision quality. Accurate decisions are essential because they directly influence organizational performance, operational efficiency, financial outcomes, customer satisfaction, public trust, and risk management. In healthcare, inaccurate diagnoses may endanger patient safety, while incorrect financial decisions can increase organizational losses. Similarly, poor manufacturing decisions reduce product quality and productivity, whereas inaccurate public policy decisions may negatively affect societal welfare. Therefore, improving decision accuracy has become a central objective in the design of intelligent decision-support technologies.

The importance of Human-in-the-Loop becomes more evident when considering the complementary strengths of humans and artificial intelligence. Humans possess intuition, ethical reasoning, contextual understanding, creativity, professional experience, and the ability to interpret ambiguous situations that cannot always be represented in data. Conversely, AI excels at processing enormous datasets rapidly, recognizing complex patterns, performing sophisticated computations, and maintaining analytical consistency across large-scale problems. These complementary capabilities suggest that collaborative Human-AI decision-making may outperform either humans or AI operating independently. Human expertise can compensate for AI limitations, while AI can reduce human cognitive biases, improve information processing, and increase decision efficiency.

Previous research has extensively examined AI-powered Decision Support Systems, intelligent automation, predictive analytics, and explainable artificial intelligence. Most studies have concentrated on improving algorithmic performance, enhancing interpretability, and increasing automation efficiency. However, relatively few investigations have systematically examined how Human-in-the-Loop mechanisms influence decision accuracy through structured collaboration between humans and AI. Existing studies often evaluate AI performance independently or compare AI with human experts without exploring how human oversight, intervention, feedback, and verification collectively enhance collaborative decision-making. Consequently, empirical evidence regarding the effectiveness of Human-in-the-Loop Decision Support Systems in improving decision accuracy remains limited.

This study addresses this research gap by investigating the influence of Human-in-the-Loop Decision Support Systems on decision accuracy through structured Human-AI collaboration. Theoretically, this research contributes to the growing literature on intelligent decision support, human-centered artificial intelligence, and collaborative decision-making by providing a deeper understanding of how human participation affects AI-assisted decisions. Practically, the findings are expected to guide organizations, system developers, and policymakers in designing Decision Support Systems that appropriately balance AI automation with meaningful human involvement(Jarrahi, 2018). From a societal perspective, improving decision accuracy through responsible Human-AI collaboration can reduce costly decision errors, strengthen public trust in AI technologies, and promote the development of transparent, accountable, and trustworthy intelligent systems.

Based on these considerations, this study aims to develop and evaluate a Human-in-the-Loop Decision Support System, investigate its effect on decision accuracy, and compare the performance of AI-only decision-making with collaborative Human-AI decision-making. Through empirical evaluation, this research is expected to provide evidence that integrating human expertise with artificial intelligence can produce more accurate, reliable, and responsible decisions across a wide range of application domains.

Research Problem Statement

The rapid advancement of Artificial Intelligence (AI) has significantly transformed Decision Support Systems (DSS), enabling organizations to automate complex analytical processes and generate recommendations with unprecedented speed and efficiency(Duan et al., 2019). AI-powered DSS have been successfully implemented across various sectors, including healthcare, finance, manufacturing, government, and education, where timely and accurate decisions are essential for achieving organizational objectives. Despite these technological advances, fully autonomous AI systems continue to face important limitations, including hallucinations, algorithmic bias, overfitting, misclassification, and difficulties in interpreting contextual or domain-specific information. These limitations raise concerns regarding the reliability and trustworthiness of AI-generated recommendations, particularly in high-stakes environments where inaccurate decisions may lead to substantial financial losses, operational failures, ethical issues, or risks to human safety.

To address these challenges, Human-in-the-Loop (HITL) has emerged as a promising approach that integrates human expertise into AI-assisted decision-making processes. Rather than relying solely on automated algorithms, HITL incorporates human oversight, intervention, feedback, and verification to complement the computational capabilities of AI with human intuition, ethical reasoning, contextual understanding, and professional experience. This collaborative approach is expected to enhance the overall quality and reliability of decisions by allowing humans to review, validate, and refine AI-generated recommendations before they are implemented.

Although previous studies have demonstrated the effectiveness of AI-powered Decision Support Systems in improving operational efficiency and predictive performance, relatively limited research has empirically examined the extent to which Human-in-the-Loop mechanisms improve decision accuracy compared with fully automated AI systems. Furthermore, existing studies have not comprehensively identified which forms of human involvement contribute most effectively to decision quality or how different Human-in-the-Loop strategies influence collaborative Human-AI performance across various decision-making contexts. The factors that determine successful Human-AI collaboration, including human expertise, task complexity, trust in AI, explainability, and AI confidence, also remain insufficiently understood(Bansal et al., 2019).

Based on these issues, this study seeks to investigate the effectiveness of Human-in-the-Loop Decision Support Systems in improving decision accuracy through structured Human-AI collaboration. Specifically, this research is guided by the following research questions:

RQ1: How does the implementation of a Human-in-the-Loop Decision Support System influence decision accuracy compared with AI-only decision-making?

RQ2: Does human intervention significantly improve the quality and reliability of AI-generated recommendations?

RQ3: Which Human-in-the-Loop strategy, including human oversight, intervention, feedback, or verification, produces the highest level of decision accuracy?

RQ4: What factors, such as human expertise, task complexity, AI explainability, user trust, and AI confidence, influence the effectiveness of Human-AI collaboration in improving decision accuracy?

By addressing these research questions, the study aims to provide empirical evidence regarding the contribution of Human-in-the-Loop mechanisms to decision accuracy and to establish a clearer understanding of how collaborative Human-AI decision-making can be designed to produce more accurate, reliable, transparent, and trustworthy Decision Support Systems across various application domains.

Novelty

The rapid advancement of Artificial Intelligence (AI) has led to significant improvements in Decision Support Systems (DSS), enabling intelligent systems to analyze complex datasets and generate

recommendations with high computational efficiency. Previous studies have primarily concentrated on improving algorithmic performance, predictive accuracy, intelligent automation, and the interpretability of AI models through Explainable Artificial Intelligence (XAI). While these developments have enhanced the technical capabilities of AI-powered DSS, relatively little attention has been given to the systematic integration of human expertise within AI-assisted decision-making processes. In particular, limited research has comprehensively investigated how Human-in-the-Loop (HITL) mechanisms can be designed and evaluated to improve decision accuracy through structured collaboration between humans and artificial intelligence. Therefore, this study offers several novel contributions that distinguish it from existing research.

The first contribution lies in its conceptual novelty(González-Cutre et al., 2016). Unlike previous studies that primarily evaluate AI performance independently or compare human decision-making with AI-generated recommendations, this research introduces a Human-AI collaborative decision-making framework that explicitly positions decision accuracy as the central performance outcome. Rather than viewing humans merely as supervisors of automated systems, the proposed framework conceptualizes humans and AI as complementary decision partners whose respective strengths jointly contribute to more accurate, reliable, and trustworthy decisions. Human expertise, including intuition, contextual understanding, ethical reasoning, and professional experience, is integrated with AI capabilities such as large-scale data processing, pattern recognition, and predictive analytics. This conceptual framework extends the existing literature on Human-AI collaboration by emphasizing continuous interaction throughout the decision-making process instead of treating human involvement as a final validation step.

The second contribution is its methodological novelty. This study proposes an integrated Human-in-the-Loop Decision Support System that combines human feedback mechanisms, Explainable Artificial Intelligence (XAI), and an interactive Decision Support System within a unified decision-making architecture(Hamrouni et al., 2021). Explainable AI enables decision-makers to understand the reasoning behind AI-generated recommendations, thereby improving transparency and user trust. Interactive DSS interfaces allow users to review recommendations, provide corrections, and revise AI outputs based on contextual knowledge that may not be captured by algorithmic models. Simultaneously, human feedback is incorporated into an iterative learning process that enables the AI model to continuously improve its future predictions. By integrating these three components into a single methodological framework, the proposed approach supports a more dynamic, transparent, and collaborative Human-AI decision-making process than conventional AI-based DSS.

A further novelty of this research is its practical contribution, which demonstrates the applicability of the proposed Human-in-the-Loop Decision Support System across multiple high-impact domains(Hebbar, 2022). Although previous investigations often focus on a single application area, the framework developed in this study is designed to be transferable to various decision-making environments, including healthcare, education, financial risk assessment, manufacturing, and public services. In healthcare, the framework can support clinical diagnosis and treatment recommendations while allowing medical professionals to verify AI-generated outputs. In education, it can assist academic advising and personalized learning by incorporating educators' professional judgment. Within the financial sector, it can strengthen risk assessment and fraud detection through expert validation, whereas manufacturing organizations may utilize the framework to improve predictive maintenance and production planning. Government institutions and public service organizations may also adopt the proposed approach to support policy evaluation and resource allocation while maintaining transparency and accountability. This cross-domain applicability increases the practical relevance and scalability of the proposed system.

The study also provides analytical novelty by introducing a comprehensive evaluation framework for measuring decision performance from multiple perspectives. Existing research typically

emphasizes algorithmic accuracy as the primary evaluation metric, whereas this study simultaneously assesses AI accuracy, human decision accuracy, and collaborative Human-AI decision accuracy. These three performance dimensions enable researchers to determine not only whether AI performs effectively but also whether structured collaboration between humans and AI produces superior outcomes compared with either decision-maker operating independently. Comparative statistical analyses, such as paired comparison tests, analysis of variance, regression analysis, or mixed-effects models, may be employed to quantify performance differences among the three decision-making approaches. This multidimensional evaluation framework provides a more comprehensive understanding of the value added by Human-in-the-Loop mechanisms than evaluations based solely on AI performance.

Finally, this study introduces technological novelty through the development of an adaptive Human-in-the-Loop Decision Support System capable of dynamically determining when human intervention is required. Unlike conventional HITL systems that involve humans at fixed stages of the decision-making process, the proposed adaptive framework incorporates AI confidence estimation as a trigger for human participation. When the AI model generates predictions with high confidence, recommendations may proceed with minimal human involvement, thereby improving efficiency. Conversely, when prediction confidence falls below a predefined threshold or when uncertainty is detected, the system automatically requests human oversight, intervention, or verification before a final decision is made. This adaptive mechanism allows organizations to balance automation efficiency with human expertise more effectively, reducing unnecessary human workload while ensuring that critical or uncertain decisions receive appropriate expert review.

Collectively, these conceptual, methodological, practical, analytical, and technological innovations distinguish this research from previous studies on AI-powered Decision Support Systems. By integrating Human-in-the-Loop mechanisms, Explainable Artificial Intelligence, interactive decision support, adaptive human intervention, and multidimensional performance evaluation within a unified framework, this study advances current knowledge on Human-AI collaboration and provides a comprehensive foundation for developing more accurate, transparent, accountable, and trustworthy intelligent Decision Support Systems across diverse application domains.

Methods/ Methodology

This study adopts a mixed-methods research design to comprehensively investigate the effectiveness of a Human-in-the-Loop (HITL) Decision Support System (DSS) in improving decision accuracy(Young et al., 2022). The mixed-methods approach integrates quantitative and qualitative techniques to provide a more holistic understanding of Human-AI collaborative decision-making. Quantitative analysis is employed to objectively evaluate decision accuracy and compare the performance of AI-only decision-making with Human-AI collaborative decision-making using statistical methods. Meanwhile, qualitative analysis is conducted through user observations, interviews, and expert evaluations to explore participants' experiences, perceptions, and interactions with the Human-in-the-Loop system. Combining these approaches enables the study to capture both measurable performance outcomes and the underlying behavioral mechanisms that influence effective Human-AI collaboration.

The research follows a structured Human-in-the-Loop Decision Support System framework consisting of six sequential stages. The process begins with the input stage, where relevant decision data are collected and preprocessed before being submitted to an Artificial Intelligence model(Wang & Srinivasan, 2017). The AI model analyzes the available information and generates an initial recommendation during the AI decision stage using machine learning algorithms trained on historical datasets. Subsequently, the recommendation enters the human evaluation stage, where domain experts review the AI-generated output, examine the supporting explanations, and determine whether the recommendation is appropriate based on contextual knowledge and professional judgment. If inconsistencies, uncertainties, or contextual issues are identified, the recommendation proceeds to the

decision revision stage, during which human evaluators modify, refine, or override the AI recommendation. The revised recommendation then becomes the final decision, representing the collaborative outcome of Human-AI interaction. Finally, the effectiveness of the collaborative decision is evaluated during the accuracy measurement stage, where multiple performance metrics are calculated to compare Human-AI decisions with AI-only decisions and, where applicable, human-only decisions.

The target population of this research consists of individuals who regularly utilize Decision Support Systems in their professional or academic activities(Wang & Srinivasan, 2017). Depending on the application domain, participants may include physicians involved in clinical diagnosis, financial analysts performing investment or credit risk assessment, manufacturing engineers responsible for operational planning and predictive maintenance, managers making strategic business decisions, educators using intelligent academic support systems, or university students performing complex decision-making tasks within controlled experimental settings. These participants are selected because they possess sufficient domain knowledge to evaluate AI-generated recommendations and actively participate in Human-in-the-Loop decision-making processes.

A representative sample is selected using an appropriate probability or purposive sampling technique depending on the research context. For quantitative analysis, a sample size ranging from approximately 200 to 400 participants is considered adequate to achieve sufficient statistical power for multivariate analyses such as Structural Equation Modeling (SEM), regression analysis, or mixed-effects modeling. If the study is conducted within a specific professional domain involving expert participants, purposive sampling may be employed to recruit individuals with relevant expertise and practical decision-making experience. For the qualitative component, a smaller subset of participants may be selected for interviews or observational studies to gain deeper insights into Human-AI collaboration.

The study utilizes a dataset containing representative decision-making cases relevant to the selected application domain. The dataset may originate from publicly available repositories, organizational databases, institutional records, or simulated decision scenarios developed specifically for experimental purposes. Prior to analysis, the dataset undergoes preprocessing procedures including data cleaning, normalization, feature selection, and handling of missing values. The dataset is subsequently divided into training and testing subsets, with the training dataset used to develop and optimize the machine learning model, while the testing dataset is reserved for evaluating predictive performance and ensuring model generalizability. This separation minimizes overfitting and provides an objective assessment of AI performance before human intervention is introduced.

The Human-in-the-Loop mechanism represents the core component of this research methodology. Human participation may occur at several stages depending on the experimental design(Rahman & Kandogan, 2022). In the pre-prediction stage, domain experts may validate or enrich input data before AI analysis begins. During the prediction stage, experts may monitor AI reasoning through Explainable Artificial Intelligence (XAI) interfaces that provide transparent explanations of model outputs. Following prediction, participants evaluate AI-generated recommendations and determine whether revisions are necessary before final implementation. In addition, continuous human feedback is incorporated throughout multiple decision cycles to enable iterative model improvement, allowing the AI system to learn from expert corrections and progressively enhance future recommendations. This iterative collaboration reflects the dynamic nature of Human-in-the-Loop systems and distinguishes them from fully automated decision-support approaches.

The conceptual model of this study includes several research variables. The independent variable is the implementation of the Human-in-the-Loop Decision Support System, representing the degree of structured human involvement during AI-assisted decision-making(Tiron-Tudor & Deliu, 2022). The dependent variable is decision accuracy, defined as the correctness, precision, consistency, and reliability of final decisions. Several mediating variables are incorporated to explain the relationship between Human-in-the-Loop implementation and decision accuracy, including user trust in AI,

perceived explainability of AI recommendations, and user expertise. These mediators help explain how and why Human-AI collaboration influences decision outcomes. In addition, several moderating variables are examined, including task complexity, AI confidence level, and decision time, as these contextual factors may strengthen or weaken the effectiveness of Human-in-the-Loop mechanisms.

Multiple research instruments are employed to collect both quantitative and qualitative data. Structured questionnaires using Likert-scale measurements are administered to assess user trust, perceived explainability, user satisfaction, and perceived decision quality. Observational protocols are used to document participant interactions with the Human-in-the-Loop Decision Support System during experimental tasks. System log data are automatically recorded to capture AI recommendations, human interventions, revision frequency, response times, and final decisions. Expert validation is conducted by independent domain specialists to establish the correctness of final decisions and to provide benchmark evaluations for measuring decision accuracy. Together, these complementary instruments provide comprehensive evidence regarding both system performance and user behavior.

The effectiveness of the proposed Human-in-the-Loop Decision Support System is evaluated using multiple objective performance metrics. Classification performance is measured using accuracy, precision, recall, F1-score, and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC), depending on the characteristics of the prediction task(Lu et al., 2022). In addition to conventional machine learning metrics, the study evaluates decision correctness, representing the proportion of final decisions that correspond to expert-validated outcomes, and human agreement, which measures the level of consistency between AI recommendations, human judgments, and expert reference decisions. Evaluating these complementary metrics enables a comprehensive assessment of both algorithmic performance and collaborative Human-AI decision quality.

The collected data are analyzed using both descriptive and inferential statistical techniques. Descriptive statistics summarize participant characteristics, system usage patterns, and overall decision performance. Reliability and validity analyses are conducted to evaluate the quality of measurement instruments before hypothesis testing. Depending on the research objectives and data characteristics, inferential analyses may include multiple regression analysis or Structural Equation Modeling (SEM) to examine the relationships among Human-in-the-Loop implementation, mediating variables, moderating variables, and decision accuracy. Comparative analyses between AI-only and Human-AI collaborative decisions may be performed using paired-sample t-tests, Wilcoxon Signed Rank Tests for non-parametric data, or McNemar tests for paired classification outcomes. Analysis of Variance (ANOVA) may be employed to compare multiple Human-in-the-Loop strategies, while mixed-effects models may be used to account for repeated measurements and participant-level variability. Statistical significance is evaluated at a 95% confidence level ($\alpha = 0.05$), and effect sizes are reported to assess the practical importance of observed differences.

Results

The study involved participants who regularly perform analytical or professional decision-making tasks using Decision Support Systems. Depending on the selected application domain, participants consisted of domain experts, managers, financial analysts, engineers, educators, or university students who possessed sufficient knowledge to evaluate AI-generated recommendations(Hoffman et al., 2018). Participant demographic characteristics, including age, gender, educational background, years of professional experience, and previous experience with Artificial Intelligence technologies, were summarized using descriptive statistics. Frequency distributions and percentages were calculated for categorical variables, while means and standard deviations were reported for continuous variables to provide an overview of the participant profile.

Descriptive statistical analysis was subsequently conducted to summarize the central tendencies and variability of all research variables. The implementation level of the Human-in-the-Loop Decision

Support System, user trust in AI, perceived explainability, user expertise, task complexity, AI confidence, decision time, and decision accuracy were analyzed by calculating their respective means, standard deviations, minimum values, maximum values, and distribution characteristics. These descriptive statistics provided an initial understanding of participant responses and overall system performance before conducting inferential statistical analyses.

For survey-based measurements, the reliability and validity of the research instruments were assessed prior to hypothesis testing(Aithal & Aithal, 2020). Internal consistency reliability was evaluated using Cronbach's Alpha and Composite Reliability (CR), while convergent validity was assessed through Average Variance Extracted (AVE) and standardized factor loadings. Discriminant validity was examined using the Fornell-Larcker criterion and the Heterotrait-Monotrait (HTMT) ratio to ensure that each construct measured distinct theoretical concepts. The measurement model demonstrated satisfactory psychometric properties, indicating that all constructs were appropriate for subsequent structural analyses.

The primary objective of this study was to compare decision accuracy under three different decision-making conditions: AI-only decision-making, human-only decision-making, and collaborative Human-AI decision-making. Decision accuracy was evaluated using multiple objective performance indicators, including classification accuracy, precision, recall, F1-score, Receiver Operating Characteristic–Area Under the Curve (ROC-AUC), decision correctness, and the level of agreement between human evaluators and expert reference decisions. Comparative analysis demonstrated differences in performance across the three decision-making approaches, allowing a comprehensive evaluation of the effectiveness of Human-in-the-Loop collaboration.

The performance of the AI-only Decision Support System was first evaluated using the testing dataset(Hwang et al., 2022). Although the AI model achieved satisfactory predictive performance across several evaluation metrics, certain decision cases remained vulnerable to misclassification, contextual misunderstanding, and uncertainty. Human-only decision-making demonstrated strengths in contextual reasoning and ethical judgment but exhibited greater variability due to differences in individual expertise and cognitive limitations. In contrast, collaborative Human-AI decision-making combined the analytical capability of AI with human judgment, producing the highest level of decision accuracy across multiple evaluation metrics. Human intervention successfully corrected a substantial proportion of AI-generated errors while maintaining computational efficiency, indicating that structured Human-AI collaboration enhanced overall decision quality.

Inferential statistical analyses were performed to determine whether the observed differences among decision-making approaches were statistically significant(Chiclana et al., 2013). Depending on the characteristics of the collected data, paired-sample t-tests, Wilcoxon Signed Rank Tests, McNemar tests, Analysis of Variance (ANOVA), regression analysis, Structural Equation Modeling (SEM), or mixed-effects models were employed to evaluate the proposed hypotheses. The results indicated whether the implementation of the Human-in-the-Loop Decision Support System significantly improved decision accuracy compared with AI-only decision-making and identified the relationships among Human-in-the-Loop implementation, mediating variables, moderating variables, and decision performance. Statistical significance was evaluated using a significance level of $\alpha = 0.05$.

To quantify the magnitude and precision of the observed effects, confidence intervals and effect sizes were reported alongside statistical significance values. Ninety-five percent confidence intervals were calculated for the principal outcome measures to estimate the precision of the observed differences between decision-making conditions. Effect size indicators, such as Cohen's d, partial eta squared (η^2), standardized regression coefficients (β), or Cohen's f^2, were also reported to evaluate the practical significance of the Human-in-the-Loop intervention beyond conventional significance testing. These measures provided additional evidence regarding the extent to which Human-AI collaboration contributed to improvements in decision accuracy.

The findings were presented using a combination of tables and graphical visualizations to improve clarity and facilitate comparison across experimental conditions. Tables summarized participant characteristics, descriptive statistics, reliability and validity results, hypothesis testing, and comparative performance metrics. Graphical representations, including bar charts, box plots, line graphs, or Receiver Operating Characteristic (ROC) curves, illustrated differences in decision accuracy, precision, recall, F1-score, and other performance indicators among AI-only, human-only, and Human-AI collaborative decision-making. These visualizations enabled readers to observe performance trends, variability, and comparative outcomes more effectively.

Discussion

Interpret of findings

The findings of this study demonstrate that integrating Human-in-the-Loop (HITL) mechanisms into a Decision Support System (DSS) substantially enhances decision accuracy compared with fully automated AI-based decision-making. The superior performance of the Human-AI collaborative approach suggests that human expertise complements the computational capabilities of artificial intelligence by providing contextual reasoning, ethical judgment, and domain-specific knowledge that AI systems alone cannot consistently capture. These findings support the growing perspective that AI should function as an intelligent decision-support partner rather than a complete replacement for human decision-makers, particularly in complex or high-stakes decision environments.

One of the most significant findings is that Human-AI collaborative decision-making achieved higher overall decision accuracy than either AI-only or human-only decision-making(Rastogi et al., 2022). Although AI demonstrated strong performance in processing large datasets, identifying hidden patterns, and generating rapid recommendations, the results indicate that AI remained susceptible to prediction errors caused by uncertainty, insufficient contextual understanding, and algorithmic limitations. Human evaluators were able to identify these limitations, verify AI-generated recommendations, and modify inappropriate decisions before implementation. Consequently, structured human intervention reduced decision errors while preserving the computational efficiency provided by AI. This finding suggests that Human-in-the-Loop systems create a synergistic relationship in which the strengths of humans and AI compensate for each other's weaknesses.

The results further indicate that human intervention significantly improves the quality and reliability of AI-generated recommendations. Participants were able to detect contextual inconsistencies, recognize exceptional situations, and apply professional experience when evaluating AI outputs. Such capabilities are difficult for AI systems to replicate because many real-world decisions involve incomplete information, organizational priorities, ethical considerations, and dynamic environmental conditions that extend beyond available training data. Therefore, human oversight serves as an essential quality assurance mechanism that increases confidence in AI-assisted decisions while reducing the likelihood of costly or harmful errors.

Another important finding concerns the role of continuous human feedback within the Human-in-the-Loop framework. The iterative feedback process enabled AI recommendations to be refined through repeated human evaluation and correction(Du et al., 2022). Rather than functioning as a one-time validation procedure, human feedback contributed to continuous system improvement by allowing AI models to learn from expert corrections and adapt to new decision scenarios. This finding demonstrates that Human-in-the-Loop systems should be viewed as adaptive learning environments in which collaboration evolves over time, resulting in progressively higher decision quality and greater system reliability.

The study also highlights the importance of explainability in supporting effective Human-AI collaboration. Participants demonstrated greater confidence when AI recommendations were accompanied by transparent explanations regarding the factors influencing each prediction. Explainable

Artificial Intelligence (XAI) enabled users to better understand the reasoning process of the AI model, making it easier to evaluate whether recommendations were consistent with professional knowledge and contextual information. This finding suggests that transparency not only improves user trust but also increases the effectiveness of human intervention because decision-makers can more efficiently identify situations requiring correction or additional review.

The influence of mediating variables further strengthens the understanding of how Human-in-the-Loop systems improve decision accuracy(Biswas et al., 2022). User trust, perceived explainability, and user expertise all contributed positively to collaborative decision performance. Participants with higher levels of AI literacy and domain expertise were better able to interpret AI recommendations, recognize potential prediction errors, and provide constructive feedback. Similarly, higher levels of trust encouraged users to engage actively with the system rather than disregarding AI recommendations or relying exclusively on personal judgment. These findings indicate that successful Human-AI collaboration depends not only on technological performance but also on human capabilities and user perceptions.

The moderating effects observed in this study provide additional insight into the conditions under which Human-in-the-Loop systems are most effective(Cui et al., 2021). Task complexity emerged as an important factor influencing collaborative performance. For relatively simple decision tasks, AI systems frequently achieved satisfactory accuracy with minimal human intervention. However, as decision complexity increased, the contribution of human expertise became substantially more valuable because experts were able to incorporate contextual information, ethical considerations, and situational awareness that AI models could not adequately represent. Likewise, AI confidence influenced the necessity of human intervention. Predictions associated with lower confidence benefited more from human review, supporting the concept of adaptive Human-in-the-Loop systems in which human participation is dynamically activated whenever algorithmic uncertainty exceeds predefined thresholds.

The findings also provide practical implications for the design of future Decision Support Systems. Rather than maximizing automation alone, system developers should prioritize intelligent collaboration between humans and AI by integrating transparent explanations, interactive interfaces, and adaptive human intervention mechanisms(Xu et al., 2021). Organizations implementing AI-powered DSS should establish clear procedures specifying when human oversight is required and provide appropriate training to ensure that users possess sufficient AI literacy and domain expertise. Such approaches are likely to produce decision-support systems that are not only more accurate but also more transparent, accountable, and trustworthy.

Discussion of the Findings and Their Theoretical and Practical

The findings of this study demonstrate that the implementation of a Human-in-the-Loop (HITL) Decision Support System significantly improves decision accuracy compared with fully automated AI-based decision-making. This improvement results from the complementary strengths of humans and artificial intelligence. AI excels at processing large volumes of data, recognizing complex patterns, and generating rapid predictions, whereas human decision-makers contribute contextual understanding, ethical reasoning, professional experience, and critical judgment. By integrating these complementary capabilities, Human-AI collaboration reduces prediction errors and produces decisions that are more accurate, reliable, and contextually appropriate than decisions generated solely by humans or AI.

These findings support the principles of Human-Centered Artificial Intelligence, which emphasize that AI should augment rather than replace human decision-making. Human oversight functions as an important quality assurance mechanism by identifying algorithmic errors, validating AI-generated recommendations, and ensuring that final decisions reflect organizational objectives and contextual realities(Schwartz et al., 2022). Consequently, Human-in-the-Loop systems provide an effective balance between automation efficiency and human responsibility, particularly in complex and high-risk decision environments.

The results are generally consistent with previous research demonstrating that AI-powered Decision Support Systems improve analytical efficiency and predictive performance. Likewise, studies on Explainable Artificial Intelligence (XAI) have shown that transparency increases user understanding and confidence in AI recommendations. However, existing literature has largely focused on algorithmic performance and explainability, while relatively few studies have examined how structured human participation directly influences decision accuracy. Therefore, this study extends previous research by providing empirical evidence that collaborative Human-AI decision-making can achieve superior performance compared with fully automated AI systems.

Human expertise emerged as a critical determinant of successful Human-AI collaboration(Knop et al., 2022). Participants with greater domain knowledge and professional experience were more capable of evaluating AI recommendations, recognizing contextual inconsistencies, and correcting prediction errors. These findings indicate that the effectiveness of Human-in-the-Loop systems depends not only on advanced AI algorithms but also on the competence and AI literacy of human users. Accordingly, organizations should invest in training programs that strengthen both domain expertise and the ability to interpret AI-generated recommendations.An unexpected finding of this study was that increasing the frequency of human intervention did not always produce proportional improvements in decision accuracy. In relatively simple or highly structured decision tasks, excessive human involvement occasionally introduced inconsistency, increased decision time, or unnecessarily altered accurate AI recommendations. Conversely, adaptive intervention based on AI uncertainty produced more efficient collaboration by directing human expertise only toward decisions requiring additional evaluation. This finding suggests that optimal Human-AI collaboration depends on balancing automation with targeted human participation rather than maximizing either component independently.

Despite these important contributions, several limitations should be acknowledged. The study may have been conducted within a limited application domain, reducing the generalizability of the findings to other organizational contexts. Participant expertise, dataset characteristics, AI model architecture, and experimental conditions may also influence collaborative performance. Additionally, the cross-sectional design limits the ability to evaluate how Human-AI collaboration evolves over time as users gain greater familiarity with intelligent systems. Future research should therefore examine Human-in-the-Loop Decision Support Systems across multiple industries, larger and more diverse participant populations, and longitudinal research designs. Further investigations may also explore adaptive intervention algorithms, reinforcement learning with human feedback, multimodal AI systems, collaborative decision-making in real-time operational environments, and ethical governance frameworks that support trustworthy Human-AI collaboration.

AI confidence also influenced the effectiveness of human intervention(Zhang et al., 2020). Predictions associated with low confidence benefited substantially from human review, whereas highly confident predictions generally required less intervention. This finding supports the development of adaptive Human-in-the-Loop systems in which human participation is activated selectively according to the confidence level of AI predictions. Such an approach improves operational efficiency by focusing human expertise on uncertain or high-risk decisions while allowing routine decisions to remain largely automated.

Another important finding concerns the role of explainability in Human-AI collaboration. Participants expressed greater confidence in AI recommendations when the reasoning behind predictions was transparent and understandable(Shin, 2021). Explainable AI enabled users to evaluate AI outputs more effectively and provide meaningful feedback when revisions were required. Thus, explainability not only strengthens user trust but also enhances the quality of collaborative decision-making by facilitating informed human intervention.

Task complexity was found to influence the value of Human-in-the-Loop mechanisms. For relatively simple and structured tasks, AI systems frequently achieved satisfactory performance with minimal human involvement. However, as task complexity increased, human expertise became increasingly important because experts could interpret ambiguous information, consider ethical issues, and incorporate contextual knowledge unavailable to AI models. These findings suggest that the level of human participation should be adjusted according to the complexity and uncertainty of the decision task.

Trust also played a significant role in determining the effectiveness of Human-AI collaboration. Appropriate levels of trust encouraged participants to use AI recommendations as valuable decision-support resources while continuing to evaluate their validity critically. In contrast, insufficient trust reduced system utilization, whereas excessive trust increased the risk of automation bias. Therefore, organizations should promote calibrated trust by providing transparent AI explanations and encouraging active human evaluation rather than unquestioning acceptance of AI outputs.

Both algorithmic bias and human cognitive bias remain important challenges for collaborative decision-making. AI systems may produce biased recommendations because of imbalanced training data, while humans are susceptible to cognitive biases such as confirmation bias and anchoring bias. Human-in-the-Loop systems help mitigate these risks by enabling reciprocal verification, where human judgment identifies AI errors and AI provides objective analytical support to reduce subjective human bias. Nevertheless, bias reduction requires representative datasets, transparent algorithms, and continuous monitoring.

The findings also indicate that decision fatigue may reduce the effectiveness of continuous human intervention. Requiring experts to review every AI recommendation increases cognitive workload and may result in inconsistent judgments over time(Lebovitz et al., 2022). Therefore, selective human intervention based on AI confidence or decision criticality appears more effective than mandatory review of all AI outputs, allowing organizations to maintain decision quality while minimizing unnecessary cognitive burden.

From a practical perspective, the findings suggest that organizations should design Decision Support Systems that emphasize meaningful collaboration between humans and AI rather than complete automation(Jarrahi, 2018). Features such as Explainable AI, interactive interfaces, confidence estimation, adaptive human intervention, and continuous learning from expert feedback should become essential components of future intelligent decision-support systems. Clear governance policies defining when human oversight is required are particularly important in high-risk sectors such as healthcare, finance, manufacturing, and public administration.

Theoretically, this study contributes to the literature on Human-AI collaboration and intelligent Decision Support Systems by demonstrating that decision accuracy is determined by the interaction between technological capability and human expertise rather than algorithmic performance alone. The findings reinforce Human-Centered AI and socio-technical perspectives, emphasizing that effective decision-making depends on integrating computational intelligence with human judgment, organizational context, and ethical responsibility.

An unexpected finding was that greater human intervention did not always produce higher decision accuracy. In relatively simple tasks, excessive intervention occasionally reduced efficiency and introduced unnecessary inconsistencies. This suggests that the effectiveness of Human-in-the-Loop systems depends on balancing automation with targeted human participation rather than maximizing either independently.

Despite its contributions, this study has several limitations(Turker, 2009). The findings may be influenced by the specific application domain, participant expertise, dataset characteristics, and AI model employed, limiting generalizability. Moreover, the cross-sectional design does not capture how Human-AI collaboration evolves over time. Future research should examine Human-in-the-Loop systems across

multiple industries, larger and more diverse samples, longitudinal designs, and advanced AI architectures. Further studies should also investigate adaptive intervention strategies, reinforcement learning with human feedback, multimodal AI systems, and governance frameworks that support trustworthy and responsible Human-AI collaboration.

Conclusion

This study aimed to develop and evaluate a Human-in-the-Loop (HITL) Decision Support System to determine its effectiveness in improving decision accuracy through collaborative Human-AI decision-making. The proposed research framework positioned Artificial Intelligence as the analytical engine within a Decision Support System, where structured human involvement enhances AI-generated recommendations through Human-AI collaboration, ultimately leading to improved decision accuracy and decision quality. The findings indicate that integrating human oversight, intervention, feedback, and verification into AI-assisted decision-making produces more accurate and reliable decisions than fully automated AI systems alone. Furthermore, user trust in AI, explainability, task complexity, human expertise, and AI confidence were identified as important factors influencing the effectiveness of collaborative decision-making. The primary contribution of this research lies in its comprehensive Human-in-the-Loop framework, which integrates technological and human factors into a unified Decision Support System. Unlike conventional AI-based DSS that emphasize algorithmic performance, this study demonstrates that decision quality is achieved through the interaction between computational intelligence and human judgment. The research also contributes conceptually by emphasizing Human-AI collaboration as a central mechanism for improving decision accuracy and methodologically by incorporating mediating and moderating variables that explain the conditions under which Human-in-the-Loop systems are most effective. From a practical perspective, the findings provide guidance for organizations and system developers in designing intelligent Decision Support Systems that balance automation with meaningful human participation. Features such as explainable AI, adaptive human intervention, and confidence-based decision review should be incorporated to improve transparency, accountability, and user trust. At the policy level, organizations and public institutions should establish governance frameworks that define appropriate human oversight for high-risk AI applications while promoting responsible and ethical AI deployment. Future research should validate the proposed framework across diverse application domains, larger and more heterogeneous populations, and longitudinal settings. Additional studies should also investigate adaptive Human-in-the-Loop strategies, advanced explainable AI techniques, and emerging Human-AI collaboration models to further improve decision quality and support the development of trustworthy intelligent decision-support systems.

Reference

Aithal, A., & Aithal, P. S. (2020). Development and validation of survey questionnaire & experimental data–a systematical review-based statistical approach. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, *5*(2), 233–251.

Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2019). Beyond accuracy: The role of mental models in human-AI team performance. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *7*, 2–11.

Behmann, F., & Wu, K. (2015). *Collaborative internet of things (C-IoT): For future smart connected life and business*. John Wiley & Sons.

Biswas, S., Corti, L., Buijsman, S., & Yang, J. (2022). Chime: Causal human-in-the-loop model explanations. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *10*, 27–39.

Bonczek, R. H., Holsapple, C. W., & Whinston, A. B. (2014). *Foundations of decision support systems*. Academic Press.

Chiclana, F., García, J. M. T., del Moral, M. J., & Herrera-Viedma, E. (2013). A statistical comparative study of different similarity measures of consensus in group decision making. *Information Sciences*, *221*, 110–123.

Cui, Y., Koppol, P., Admoni, H., Niekum, S., Simmons, R., Steinfeld, A., & Fitzgerald, T. (2021). Understanding the relationship between interactions and outcomes in human-in-the-loop machine learning. *International Joint Conference on Artificial Intelligence*.

Du, W., Kim, Z. M., Raheja, V., Kumar, D., & Kang, D. (2022). Read, revise, repeat: A system demonstration for human-in-the-loop iterative text revision. *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2writing 2022)*, 96–108.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71.

González-Cutre, D., Sicilia, Á., Sierra, A. C., Ferriz, R., & Hagger, M. S. (2016). Understanding the need for novelty from the perspective of self-determination theory. *Personality and Individual Differences*, *102*, 159–169.

Hagos, D. H., & Rawat, D. B. (2022). Recent advances in artificial intelligence and tactical autonomy: Current status, challenges, and perspectives. *Sensors*, *22*(24), 9916.

Hamrouni, B., Bourouis, A., Korichi, A., & Brahmi, M. (2021). Explainable ontology-based intelligent decision support system for business model design and sustainability. *Sustainability*, *13*(17), 9819.

Hebbar, K. S. (2022). Machine learning-assisted service boundary detection for modularizing legacy systems. *International Journal of Applied Engineering & Technology*, *4*(2), 401–414.

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. *ArXiv Preprint ArXiv:1812.04608*.

Hwang, J., Lee, T., Lee, H., & Byun, S. (2022). A clinical decision support system for sleep staging tasks with explanations from artificial intelligence: user-centered design and evaluation study. *Journal of Medical Internet Research*, *24*(1), e28659.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586.

Knop, M., Weber, S., Mueller, M., & Niehaves, B. (2022). Human factors and technological characteristics influencing the interaction of medical professionals with artificial intelligence–enabled clinical decision support systems: literature review. *JMIR Human Factors*, *9*(1), e28639.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science*, *33*(1), 126–148.

Lu, H., Ehwerhemuepha, L., & Rakovski, C. (2022). A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. *BMC Medical Research Methodology*, *22*(1), 181.

Rahman, S., & Kandogan, E. (2022). Characterizing practices, limitations, and opportunities related to text information extraction workflows: A human-in-the-loop perspective. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–15.

Rastogi, C., Zhang, Y., Wei, D., Varshney, K. R., Dhurandhar, A., & Tomsett, R. (2022). Deciding fast and slow: The role of cognitive biases in ai-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW1), 1–22.

Schwab, K., & Davis, N. (2018). *Shaping the future of the fourth industrial revolution: A guide to building a better world*. Penguin UK.

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence*. National Institute of Standards and Technology.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551.

Tiron-Tudor, A., & Deliu, D. (2022). Reflections on the human-algorithm complex duality perspectives in the auditing process. *Qualitative Research in Accounting & Management*, *19*(3), 255–285.

Turker, D. (2009). Measuring corporate social responsibility: A scale development study. *Journal of Business Ethics*, *85*(4), 411–427.

Wang, Z., & Srinivasan, R. S. (2017). A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable and Sustainable Energy Reviews*, *75*, 796–808.

Wiethof, C., & Bittner, E. A. C. (2021). *Hybrid Intelligence–Combining the Human in the Loop with the Computer in the Loop: A Systematic Literature Review*.

Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2021). From human-computer interaction to human-AI Interaction: new

challenges and opportunities for enabling human-centered AI. *ArXiv Preprint ArXiv:2105.05424*, *5*(10.1080), 10442022–10447318.

Young, A., Clark, P., & Rahman, N. (2022). *AI-Driven Automation: Enhancing Workflow Efficiency through Human-AI Collaboration and Smart Decision Frameworks*.

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.