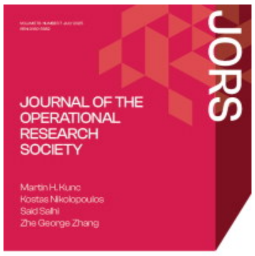


Journal of the Operational Research Society

ISSN: 0160-5682 (Print) 1476-9360 (Online) Journal homepage: www.tandfonline.com/journals/tjor20

Human-AI coordination for large-scale group decision making with heterogeneous feedback strategies

Jing Zhang, Ning Wang & Ming Tang

JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY
https://doi.org/10.1080/01605682.2025.2466677

ORIGINAL ARTICLE

Human-AI coordination for large-scale group decision making with heterogeneous feedback strategies

Jing Zhang, Ning Wang and Ming Tang

School of Economics and Management, Dalian University of Technology, Dalian, China

ABSTRACT

In group decision making, human experts are usually susceptible to cognitive biases and information overload. Artificial intelligence (AI) has capabilities in data processing and analysis, but is limited by issues such as interpretability and human adoption. Humans and AI have different problem-solving capabilities, they can benefit from each other. Thus, there is a need to leverage a mechanism to tap into the intelligence of both parties and achieve mutually shared outcomes. In this study, we propose a large-scale group decision-making model with human-AI consensus. First, an improved density-peak clustering algorithm is utilized to classify experts into subgroups based on the Similarity-Trust-Attitude score. Then, weights of experts and subgroups are obtained based on the internal influence of experts and the intuitionistic fuzzy entropy of subgroup preferences. Further, considering three different strategies of human-AI interaction, subgroup consensus and subgroup-AI consensus are calculated. Finally, a minimum cost consensus model with heterogeneous feedback strategies is proposed. The usability of the proposed model is verified through a medical diagnosis case. This study found that the human-AI coordination with heterogeneous feedback strategies can reduce adjustment costs, and different interaction mechanisms have different effects.

ARTICLE HISTORY
Received 5 June 2024
Accepted 7 February 2025

KEYWORDS
Human-AI coordination; large-scale group decision making; minimum cost consensus; subgroup; medical diagnosis

1. Introduction

Group decision making (GDM) aims to facilitate the optimal ranking of a set of alternatives to a specific problem through the collaborative effort of multiple experts (Meng et al., 2024). With the increasing complexity of decision-making settings and the rising demand for cross-disciplinary expertise to tackle decision-making problems, large-scale group decision making (LSGDM) has become a focal point of research (Tang & Liao, 2021). To fully leverage the role of a large group, scholars have proposed many models from various perspectives to extract and harness human intelligence, such as expert behaviours (Gou et al., 2021), subgroup classification (Liang et al., 2024; Tang et al., 2020), consensus-building method (Meng et al., 2023) and preference information aggregation (Chen et al., 2022). Thus, traditional LSGDM relies mostly on the experience and intuition of human experts. However, human experts may be influenced by subjective biases (Liang et al., 2024), information overload (Laker et al., 2018), and limited learning capacity (Boyacı et al., 2024), resulting in uncertainty and bias in decision-making outcomes.

Nowadays, with the development in computational power of information technologies and data availability, scholars and business organizations are no longer limited to using human intelligence and have begun to utilize artificial intelligence (AI) techniques to develop a variety of GDM systems in a wide range of fields (Garcia et al., 2024). For example, AI helps clinicians make clinical decisions in vascular surgery, including patient classification, early diagnosis, and surgical risk assessment, by analyzing huge and unstructured data (Lareyre et al., 2023). Finnish IT company Tieto announced the appointment of AI to their leadership team (Zheng et al., 2023). Additionally, Generative Artificial Intelligence (GenAI), as an emerging technology in the field of AI, is increasingly demonstrating its ability to make advantageous predictions (Hermann & Puntoni, 2024). Understanding features and effective tips of GenAI can help people make wise decisions regarding their problems (Beilby & Hammarberg, 2024). For example, in the context of medical decision making, Sorin et al. (2023) explored the utilization of ChatGPT as a decision support tool for breast tumour boards. They provided ChatGPT-3.5 with clinical information about 10 consecutive patients. They then asked the chatbot to perform three tasks: summarization, recommendation, and explanation. In 70% of cases, the chatbot's recommendations agreed with the tumour

CONTACT Ming Tang tangming@dlut.edu.cn School of Economics and Management, Dalian University of Technology, Dalian 116024, China.

board's decisions. However, Saban and Dubovi (2024) found that GenAI tools exhibited a proclivity towards indecision and over-triage in comparison to human clinicians. Ray et al. (2023) argued that GenAI provides clinicians with powerful tools to improve surgical precision, diagnostic accuracy, and treatment planning, but AI was meant to flag relevant studies for human review, rather than replacing the physician's assessment.

While scholars acknowledge the potential of GenAI in GDM, it still has shortcomings compared to human intelligence such as limited interpretability, the possibility of generating false information, restricted reasoning capabilities, and ethical concerns (Ray et al., 2023; Saban & Dubovi, 2024; Thuy & Benoit, 2024; Vasey et al., 2022). Human experts utilize cognitive flexibility to integrate various sources of information, but they are limited by their cognitive capacity when it comes to evaluating large volumes of data. Fortunately, both human and AI have their own strengths, possessing the potential to complement each other's shortcomings. A nascent agreement is to employ the symbiotic relationship between human and AI (Wang et al., 2024). For example, the coordination between human radiologists and AI improved the overall accuracy of pneumonia diagnosis compared to radiologists or AI alone (Patel et al., 2019). Given the strengths of both human experts and AI, researchers are increasingly acknowledging the benefits of reciprocal learning—a form of human-AI learning loop that leverages the complementary capabilities of each to achieve consensus (Jussupow et al., 2021; Sturm et al., 2021). Therefore, it is critical to combine strengths of AI and human intuition to ensure the continued viability of decision-making processes (Choudhary et al., 2025; Te'eni et al., 2023).

To fully harness the strengths of both human experts and AI in the consensus, it is necessary to develop a LSGDM model considering heterogeneous preferences of experts and AI. Consensus building-approach based on multiple feedbacks is a way to achieve a shared situation among different stakeholders. Consensus-based human-AI decision-making model can promote deep integration of LSGDM processes. However, addressing this problem entails navigating several critical challenges. The first challenge is the measurement of GenAI's participation degree in LSGDM. In traditional LSGDM, the weights of experts are often determined based on the trustworthiness of experts in the group. However, in real life, due to factors such as humans' trust and acceptance on AI, AI's decision in different scenarios may be endowed with varying degrees of participation. The second is the measurement of the consensus between experts and GenAI. Consensus is traditionally measured based on whether the consensus of a group exceeds a preset threshold. If GenAI is integrated into the group, how to measure the reliability of GenAI's decision and effectively induce the consensus between GenAI and experts is an important issue. The third is the determination of the feedback strategy for the consensus-building approach. In traditional LSGDM, feedback is generated by selecting experts or subgroups of experts who are not reached the consensus threshold and providing suggestions for them to revise their preferences. However, in feedback strategies involving GenAI, experts may not have sufficient willingness to cooperate with GenAI.

To address these above challenges, we propose an LSGDM model to combine GenAI and experts' preferences, and develop a consensus-building approach. We investigate the unity of subgroup consensus and subgroup-AI consensus. The impact of GenAI on the consensus-building process will be explored.

The main contributions of our research are threefold: (1) The impact of GenAI's different levels of participation on the consensus process is studied. Feedback strategies are conducted based on three different scenarios (GenAI as the decisive role, GenAI as the equal role, and GenAI as the auxiliary role). (2) A threshold measure based on subgroup consensus and subgroup-AI consensus is proposed. We introduce a new weight measurement method to determine the standard weight with different decision-making bodies, and determine the consensus threshold based on subgroup consensus and subgroup-AI consensus. (3) A minimum cost consensus (MCC) model based on heterogeneous feedback strategies is constructed. Feedback adjustment parameters of experts are set based on experts' attitudes towards GenAI. Finally, the applicability of our proposed model is verified based on a medical diagnosis case.

The remainder of this paper is organized as follows. Section 2 is a literature review, including human-AI decision making, LSGDM and MCC models. Section 3 proposes the main model with heterogeneous feedback strategies in three scenarios. Section 4 illustrates application of our proposed model using a case study about medical diagnosis. Section 5 presents comparative analysis. Section 6 concludes the paper with implications, limitations and future work.

2. Literature review

In this section, we review two streams of relevant literature: (1) Human-AI decision making, (2) LSGDM and MCC models.

2.1. Human-AI decision making

AI, which refers to "the ability of a system to identify, interpret, make inferences, and learn from data to achieve predetermined organizational and societal goals" (Mikalef & Gupta, 2021) has been incorporated into algorithms to make them more powerful (Mahmud et al., 2022). However, there are still ethical issues associated with the use of AI, such as data privacy and the potential for AI bias to influence the outcome of decisions. Therefore, people have different attitudes toward the use of AI, and these attitudes also directly affect the final perception of experts' decisions (Mahmud et al., 2024). People embracing AI believe that human-AI coordination can accelerate decision-making efficiency (Mahmud et al., 2024). People with AI aversion may worry about negative impacts of AI (Burton et al., 2020).

Certainty, the most promising application of AI is not to replace humans, but rather synergize with them, creating "superminds" capable of accomplishing cognitive and physical tasks that were once considered impossible (Malone et al., 2020). An increasing number of studies have found that when humans and AI make decisions together, the overall quality of decisions is improved (Patel et al., 2019). However, the effect of AI's decision varies across domains. For example, Selby et al. (2024) found that Large Language Model based inference has good downstream performance in medicine, business, and biology, but not in other domains. Table 1 reviews related literature on human-AI decision making. We found that AI has been assigned different roles in human-AI teams.

2.2. LSGDM and MCC models

LSGDM is a process in which a large number of experts evaluate a set of feasible alternatives and select the best solution. Since there are many stakeholders with different backgrounds in LSGDM, consensus reaching-process (CRP) is particularly important in LSGDM. CRP is an iterative and dynamic process. In this process, experts discuss and modify their initial preferences to reach a collective preference that is satisfactory to all experts. Such a process is usually guided and supervised by a person called the supervisor (Tang et al., 2020). The typical consensus model of LSGDM includes four aspects: preference aggregation, degree of consensus measurement, feedback mechanism design and alternative selection.

To build a CRP at the smallest possible cost, Ben-Arieh and Easton (2007) first introduced the concept of MCC model. Later, many scholars made improvements on this model. Table 2 presents a summary of some MCC models in LSGDM. Existing models set experts' adjustment cost according to preference gap, or use the same adjustment cost. However, this situation ignores acceptance of preference adjustment by experts. Therefore, to make up for the existing limitations, we consider experts' attitudes toward GenAI when adjusting

Table 1. Literature related to human-AI decision making.

Article	Human-AI interaction type	Research methods	Application field
(Boyacı et al., 2024)	AI-supported human decision making	Decision model based on the framework of cognitive flexibility and limited capacity	Medical diagnostics
(Garcia et al., 2024)	AI-supported human decision making	Parsimonious model of advice	Hotel pricing
(Saban & Dubovi, 2024)	AI-supported human decision making	Experimental investigation	Clinical decisions
(Zheng et al., 2023)	Equal decision-making power	A wizard-of-oz study	English articles review and rank
(De Véricourt et al., 2023)	AI performs decision tasks under human supervision	Repeated tasks and learning	Decide whether a specific action (e.g., a biopsy)
(Ge et al., 2021)	AI-supported human decision making	Empirical study	Financial-advising services in peer-to-peer lending

Table 2. Literature on MCC in LSGDM.

Article	MCC categories	Unit adjustment cost	Feedback	Model characteristics
(Zhang et al., 2022)	MCC in heterogeneous opinion groups	Heterogeneous	No	MCC with large-scale heterogeneous group decision making
(Liang et al., 2023)	MCC with loss aversion	Heterogeneous	Yes	Two-stage consensus reaching mechanism
(Yu et al., 2022)	MCC in social network LSGDM	Heterogeneous	Yes	MCC considering voluntary trust loss
(Liang et al., 2022)	Minimum cost of informed individuals and time constraint	Heterogeneous	Yes	LSGDM with bounded confidence effects
(Du et al., 2024)	MCC in social network LSGDM	Heterogeneous	Yes	Constrained community detection method and a multistage multicost consensus
(Li et al., 2023)	MCC in dynamic social networks LSGDM	Homogenous	Yes	Two-stage consensus reaching mechanism

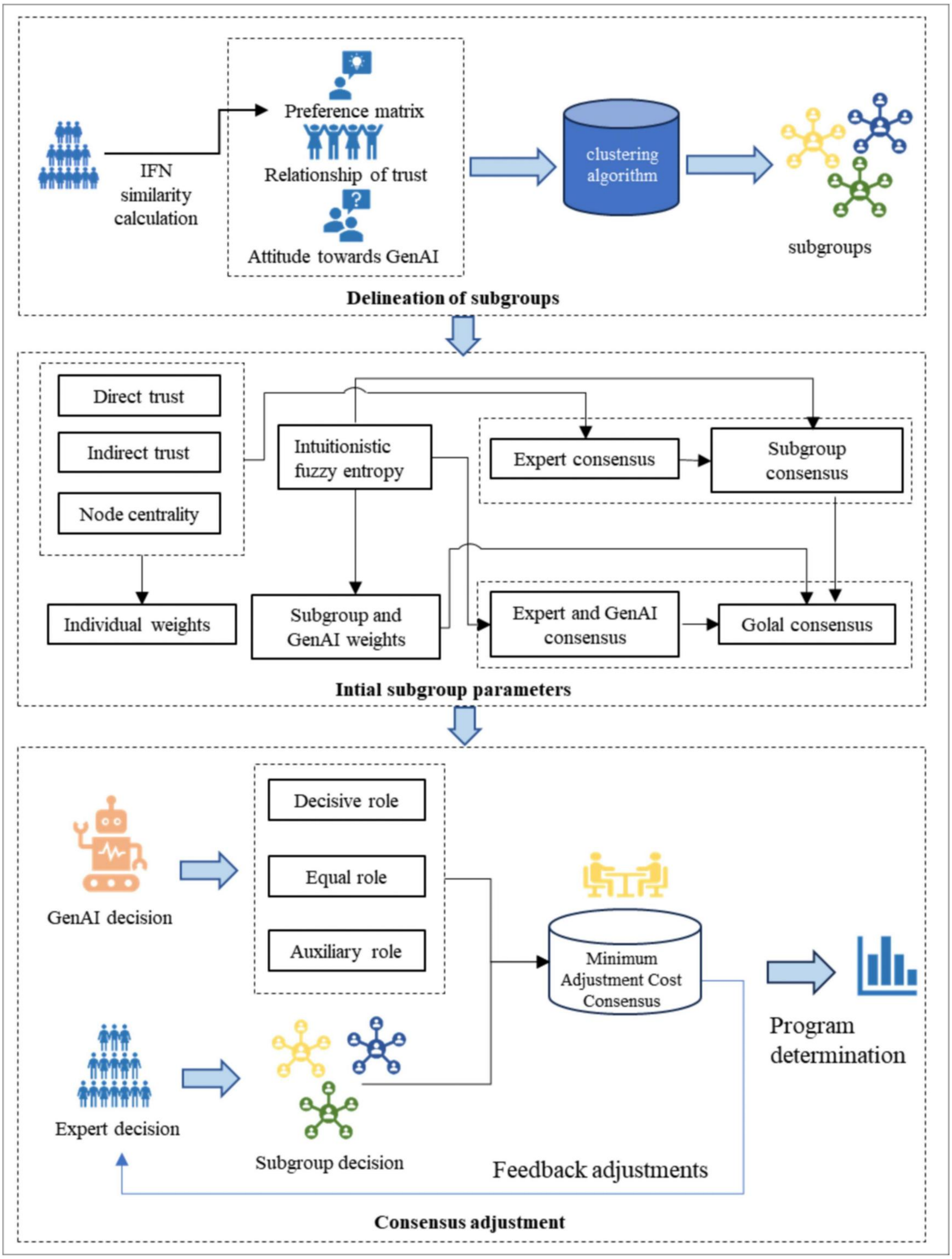


Figure 1. CRP of our model.

preferences, and use a MCC model to perform heterogeneous feedback strategies.

3. LSGDM model based on human-AI coordination

3.1. Model setup

The CRP process of our model is shown in Figure 1. We develop a CRP to achieve an outcome between experts and GenAI. First, the model uses an improved Density Peaks Clustering (DPC) algorithm to classify large-scale experts into subgroups. Then, weights of experts, subgroups and GenAI are introduced. On this basis, considering three levels of participation of GenAI, the degrees of consensus at different levels are calculated, and a MCC model with heterogeneous feedback strategies is proposed. Finally, the model is applied to a medical diagnosis example to verify its feasibility.

The LSGDM includes the following components:

a. A fixed set of feasible alternatives $X = \{x_1, x_2, \ldots, x_m\}, m \geq 2$.

Table 3. Notations of this study.

Notation	Explanation
$E = \{e_1, e_2, ..., e_q\}$	The set of experts
$X = \{x_1, x_2, ..., x_m\}$	The set of alternatives
$C = \{c_1, c_2, ..., c_n\}$	The set of criteria
$Z = (z_1, z_2, ..., z_n)^T$	The weight vector of criteria
$P^k = (p_{ij}^k)_{m \times n}$	Decision-making matrix of expert e_k
$P^{AI} = (p_{ij}^{AI})_{m \times n}$	Decision-making matrix of GenAI
$P^{ave} = (p_{ij}^{ave})_{m \times n}$	Average decision-making matrix of all subgroups
$P^{DAI} = (p_{ij}^{DAI})_{m \times n}$	Average decision-making of GenAI-experts
C_l	$l - th$ subgroup
w_l^k	The weight of e_k in subgroup C_l
CD_l^k	Individual consensus degree
ACD_l	Consensus degree of subgroup
$AACD_{l-AI}$	Consensus degree of subgroup and GenAI
$GACD$	Global consensus
σ	Consensus threshold
c_k	Unit cost for adjusting expert e_k's preference
g_{ij}	The final global preference information of alternative x_i regarding criterion c_j
G_i	The decision-making information of alternative x_i

b. A group of experts $E = \{e_1, e_2, ..., e_q\}$ who expresses their preferences on X.

c. A fixed set of criteria $C = \{c_1, c_2, ..., c_n\}$. The weight vector of criteria is $Z = (z_1, z_2, ..., z_n)^T$, where $z_j \geq 0$ and $\sum_{j=1}^{n} z_j = 1$.

d. Let $P^k = (p_{ij}^k)_{m \times n} (k = 1, 2, ..., q)$ be decision-making matrix composed of intuitionistic fuzzy numbers (IFNs) given by $e_k \in E$, where p_{ij}^k denotes the preference of e_k for x_i regarding c_j. Then, the intuitionistic fuzzy set (IFS) A on X is defined as

$$A = \{x, \mu_A(x), \nu_A(x) | x \in X\},$$

where $\mu_A : X \to [0, 1], x \in X \to \mu_A(x) \in [0, 1]$ and $\nu_A : X \to [0, 1],\ x \in X \to \nu_A(x) \in [0, 1]$ respectively denote the membership and non-membership degrees of element $x \in X$ to A, and $0 \leq \mu_A(x) + \nu_A(x) \leq 1, x \in X$. Furthermore, the function $\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$ is the hesitation degree of x to A.

With the rapid advancement of AI technology, AI has been able to evaluate the reliability of the prediction results. For example, the AI model designed by Ma et al. (2023) can give the classification outcome of the task prediction and the confidence of AI. Thus, this study adopts IFS to represent GenAI's preference. Let $P^{AI} = (p_{ij}^{AI})_{m \times n}$ be GenAI's decision-making matrix, where p_{ij}^{AI} denotes the preference of GenAI for x_i regarding c_j.

e. There exists a social network based on cooperative relationships among experts, which is represented as an undirected weighted graph $G = (E, L)$, where E denotes the set of nodes, L denotes the set of edges between nodes. In the social network, edges represent trust connections between experts and nodes represent experts. And let $ST = (St_{kh})_{q \times q}$ be the trust matrix among experts. Furthermore, $SA = \{sa_1, sa_2, ..., sa_q\}$ denotes experts' attitudes towards GenAI, $0 \leq sa_k \leq 1$.

For ease of understanding, Table 3 gives the main symbols used in this paper and their corresponding descriptions.

3.2. Group clustering

Let $A = \{\langle x, \mu_A(x), \nu_A(x)\rangle | x \in X\}$ and $B = \{\langle x, \mu_B(x), \nu_B(x)\rangle | x \in X\}$ be two IFSs. We use the normalized Hamming distance measure to calculate the distance between A and B (Xu, 2007):

$$D(A, B) = \frac{1}{2n} \sum_{i=1}^{n} (|\mu_A(x_i) - \mu_B(x_i)| + |\nu_A(x_i) - \nu_B(x_i)| + |\pi_A(x_i) - \pi_B(x_i)|) \quad (1)$$

$$Sim(A, B) = 1 - D(A, B) \quad (2)$$

Based on Eq. (1), the distance between two experts can be obtained; based on Eq. (2), experts' similarity matrices can be obtained.

To assess the cohesion of subgroups, we combine experts' preference similarity, social trust and attitudes towards GenAI to obtain the Similarity-Trust-Attitude (STA) score. We assume that preference similarity, trust relationship and attitude are equally important in clustering process, and we draw Einstein's T-paradigm idea to calculate the STA score.

Let STA_{kh} be the STA score from expert e_k to e_h, which can be obtained by:

$$STA_{kh} = \frac{Sim_{kh} \cdot St_{kh} \cdot Sa_{kh}}{1 + (1 - Sim_{kh})(1 - St_{kh})(1 - Sa_{kh})} \quad k, h = 1, 2, ..., q \quad (3)$$

where Sim_{kh} is the similarity between e_k and e_h, St_{kh} is the trust degree from e_k to e_h, $Sa_{kh} = 1 -$

$|sa_k - sa_h|$ is the GenAI-Attitude similarity between e_k and e_h. In Eq. (3), $0 \leq STA_{kh} \leq 1$, $STA_{kh} \leq \min\{Sim_{kh}, St_{kh}, Sa_{kh}\}$. The higher the value of STA_{kh}, the higher the probability that e_k and e_k are classified into the same subgroup.

The DPC algorithm has been widely used in various data clustering tasks (Li et al., 2018; 2024). Unlike algorithms such as k-means, which require a pre-specified number of clusters, the DPC algorithm can recognize data of any shape, visually identify the number of clusters in the data and effectively identify outliers (Meng et al., 2023). However, the stability of the clustering centre selection in the DPC algorithm is relatively low, in this study, we proposed an improved DPC algorithm to enhance the stability of clustering centres and applied it to classify experts. Algorithm 1 presents the steps of the improved DPC algorithm.

Algorithm 1. Pseudo-code for the improved DPC algorithm.

Input: STA matrix of experts, parameter values d_c, α, θ_1, η_θ
Output: Clustering results
Step 1: Calculate the distance matrix D among experts based on the STA matrix of experts.
Step 2: Determine the intercept distance d_c
Step 3: Compute the local density ρ based on the formula $\rho_i = \sum_{i \neq j} \exp\left(-(d_{ij}/d_c)^2\right)$.

Step 4: Compute the relative distance δ based on the formula
$$\delta_i = \begin{cases} \min_{j:\rho_j > \rho_i} d_{ij} & \text{if } \rho_i < \max(\rho) \\ \max_{j \neq i} d_{ij} & \text{if } \rho_i = \max(\rho) \end{cases}$$
Step 5: Select stable cluster centres:
for i in n do
calculate the decision value $r_i = \rho_i \cdot \delta_i$
calculate the density of data nearest 5 neighbors $\{\rho_j : j \in \varsigma_i\}$ //ς_i is the set of neighboring points
evaluate stability: $\eta_i = \alpha \cdot \rho_i + (1-\alpha)\delta_i / std(\varsigma_i)$ //std is the standard deviation
centres satisfies $r_i > r_\theta$ and $\eta_i > \eta_\theta$ //$r_\theta = percentile(r, \theta_1)$
end for
Step 6: Assign each point directly to the nearest cluster centre based on the distance matrix.
Step 7: Output the clustering results.

3.3. Weight determination and consensus measure

3.3.1. Weight determination

After clustering, the weights of experts in subgroups and weights of subgroups should be determined.

Step 1. Determine the weights of experts in a subgroup.

The weight of an expert in a group is calculated based on the combination of direct trust, indirect trust, and expert centrality:

$$w_l^k = \omega_1 \cdot DT_k + \omega_2 \cdot IT_k + \omega_3 \cdot CT_k \tag{4}$$

where w_l^k denotes the weight of e_k in C_l, DT_k denotes the direct trust of e_k, IT_k denotes the indirect trust of e_k, and CT_k denotes the degree centrality of e_k. ω_1, ω_2 and ω_3 corresponds to initial weight of three items in Eq. (4), respectively. $\omega_1 + \omega_2 + \omega_3 = 1$, $\omega_1, \omega_2, \omega_3 > 0$.

Step 2. Calculate the weights of subgroups and GenAI.

In general, the greater the self-confidence of subgroup members, the less they are affected by social influence, and the more likely they adhere to their initial preferences (Tu et al., 2024). According to the characteristic of IFN, a smaller intuitionistic fuzzy entropy means a greater evaluation confidence. Therefore, we calculate the weights of subgroups and GenAI by the intuitionistic fuzzy entropy.

According to Chen et al. (2022), for an IFS $A = \{x, \mu_A(x_i), \nu_A(x_i) | x_i \in X\}$, the intuitionistic fuzzy entropy of A is defined as:

$$E(A) = \frac{1}{n}\sum_{i=1}^{n} \frac{(1-\Delta_A(x_i))e^{(1-\Delta_A(x_i))} + \pi_A(x_i) \cdot e^{\pi_A(x_i)}}{e^{(1-\Delta_A(x_i))} + e^{\pi_A(x_i)}} \tag{5}$$

where $\Delta_A(x_i) = |(\mu_A(x_i) - \nu_A(x_i))/(\mu_A(x_i) + \nu_A(x_i))|$ represents the power of IFS A.

Based on the above definition of intuitionistic fuzzy entropy, the subgroup average decision-making matrix $P^l = (p_{ij}^l)_{m \times n}$ and the GenAI decision-making matrix $P^{AI} = (p_{ij}^{AI})_{m \times n}$ are constructed. If only the proportion of weights among subgroups is considered, the weight of a subgroup can be obtained as:

$$w_l = \frac{1 - \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n} z_j \cdot E(p_{ij}^l)}{L - \frac{1}{m}\sum_{l=1}^{L}\sum_{i=1}^{m}\sum_{j=1}^{n} z_j \cdot E(p_{ij}^l)} \tag{6}$$

If the proportion of weights among subgroups and GenAI is considered, the standardized weight of a subgroup or GenAI is

$$w_s = \frac{1 - \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n} z_j \cdot E(p_{ij}^s)}{S - \frac{1}{m}\sum_{s=1}^{S}\sum_{i=1}^{m}\sum_{j=1}^{n} z_j \cdot E(p_{ij}^s)} \tag{7}$$

where w_s denotes the weight of a subgroup or GenAI, L is the number of subgroups, S is the sum of number of subgroups and GenAI, where $p_{ij}^s = < \mu_{ij}^s, \nu_{ij}^s >$ denotes the preference of subgroup or GenAI for x_i under c_j.

3.3.2. Consensus measure

In this section, we calculate the degree of consensus at expert level, subgroup level, subgroup-AI level and global level, respectively.

Expert level: the degree of consensus among experts within a subgroup.

For e_k in C_l, the degree of individual consensus CD_l^k is a measure of the proximity between e_k's decision-making matrix $P^k = (p_{ij}^k)_{m\times n}$ and subgroup's average decision-making matrix $P^l = (p_{ij}^l)_{m\times n}$. CD_l^k is defined as:

$$CD_l^k = 1 - d(P^k, P^l) = 1 - \frac{1}{2m}\sum_{i=1}^{m}\sum_{j=1}^{n} z_j \left(\left|\mu_{ij}^k - \mu_{ij}^l\right| + \left|\nu_{ij}^k - \nu_{ij}^l\right| + \left|\pi_{ij}^k - \pi_{ij}^l\right|\right) \tag{8}$$

Subgroup level: the degree of consensus among subgroups (without AI).

For C_l, the subgroup consensus ACD_l is a measure of the proximity between subgroup's average decision-making matrix $P^l = (p_{ij}^l)_{m\times n}$ and average decision-making matrix $P^{ave} = (p_{ij}^{ave})_{m\times n}$ of all subgroups. ACD_l is defined as:

$$ACD_l = 1 - d(P^{ave}, P^l) = 1 - \frac{1}{2m}\sum_{i=1}^{m}\sum_{j=1}^{n} z_j \left(\left|\mu_{ij}^{ave} - \mu_{ij}^l\right| + \left|\nu_{ij}^{ave} - \nu_{ij}^l\right| + \left|\pi_{ij}^{ave} - \pi_{ij}^l\right|\right) \tag{9}$$

Subgroup-AI level: the degree of consensus between subgroups and GenAI.

The degree of consensus $AACD_{l-AI}$ is the measure of the proximity between subgroup's average decision-making matrix $P^l = (p_{ij}^l)_{m\times n}$ and GenAI's decision-making matrix $P^{AI} = (p_{ij}^{AI})_{m\times n}$. $AACD_{l-AI}$ is defined as:

$$AACD_{l-AI} = 1 - d(P^{AI}, P^l) = 1 - \frac{1}{2m}\sum_{i=1}^{m}\sum_{j=1}^{n} z_j \left(\left|\mu_{ij}^{AI} - \mu_{ij}^l\right| + \left|\nu_{ij}^{AI} - \nu_{ij}^l\right| + \left|\pi_{ij}^{AI} - \pi_{ij}^l\right|\right) \tag{10}$$

Global level: the global consensus for the whole group

The degree of global consensus $GACD$ is a measure of the final degree of consensus among experts and GenAI. We use the minimum value of ACD_l and $AACD_{l-AI}$ to represent $GACD$, which allows us to control the threshold for both subgroup consensus and subgroup-AI consensus. Then, the global consensus level $GACD$ can be obtained as:

$$GACD = min\{ACD_l, AACD_{l-AI}\} \tag{11}$$

It is easy to prove $0 \le GACD \le 1$. A larger value of $GACD$ indicates a higher degree of consensus. $GACD=1$ indicates that all subgroups and GenAI have reached a unanimous consensus. It is assumed that $\sigma (0 < \sigma < 1)$ is a predefined group consensus threshold. If $GACD \ge \sigma$, group consensus is considered to be reached; if $GACD < \sigma$, it means that the degree of decision consensus has not yet been reached and experts' preferences or GenAI's preference need to be adjusted to achieve the threshold.

3.4. Heterogeneous feedback strategies and selection process

3.4.1. Feedback scenarios

If the group consensus threshold is not reached, it is necessary to provide feedback to adjust experts' preferences or GenAI's preference and then recalculate the degree of consensus until the threshold is reached. Therefore, we propose a MCC model based on experts' attitudes towards GenAI and give different feedback strategies according to human-AI interaction scenarios. To better reflect the interaction between GenAI's decision and human experts' decisions, GenAI also takes part in each round of consensus iteration.

1. When GenAI's decision and the average decision of subgroups is consistent (the degree of consensus is within the threshold), GenAI will not change its decision;
2. When GenAI's decision and the average decision of subgroups is inconsistent (the degree of consensus is not within threshold), GenAI will modify its decision towards the average decision of subgroups.

The final modification suggestion $p_{ij}^{AI'}$ will be generated by combining the initial preference p_{ij}^{AI} and the modification preference p_{ij}^{ave}. The updated formula is calculated as (Li et al., 2023):

$$p_{ij}^{AI'} = IFWA\left\langle p_{ij}^{AI}, p_{ij}^{ave}\right\rangle = \left\langle 1-\left(1-\mu_{ij}^{AI}\right)^{\vartheta}\cdot\left(1-\mu_{ij}^{ave}\right)^{1-\vartheta}, \left(\nu_{ij}^{AI}\right)^{\vartheta}\cdot\left(\nu_{ij}^{ave}\right)^{1-\vartheta}\right\rangle \tag{12}$$

where p_{ij}^{AI} denotes the preference of GenAI that needs feedback adjustment, and p_{ij}^{ave} denotes the average decision of all subgroups. $\vartheta \in [0,1]$ is an adjustment parameter, which indicates the degree of retention of GenAI to its own preference.

The following focuses on analyzing the feedback performance of experts in different scenarios.

3.4.1.1. Scenario I: GenAI as the decisive role.

The importance of GenAI's decision in the decision-making process lies in its ability to handle large-scale data analysis or realize rapid response, etc. For example, in areas such as climate prediction and high-frequency trading, GenAI not only improves

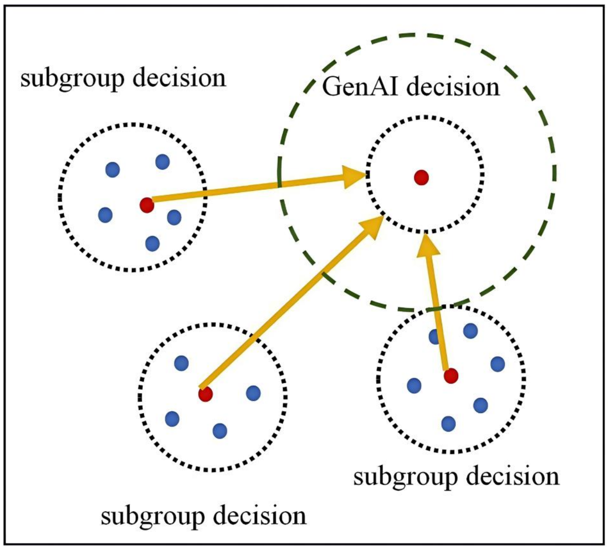


Figure 2. Adjustment direction of subgroups' decisions in Scenario I.

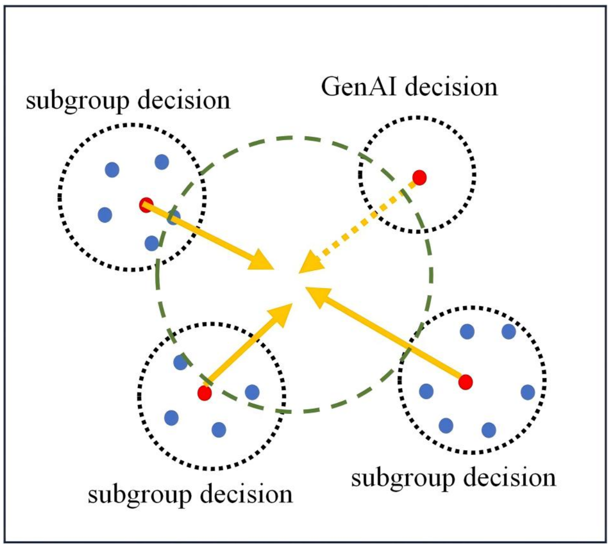


Figure 3. Adjustment direction of subgroups' decisions in Scenario II.

the speed and accuracy of decisions, but also enable effective decisions in areas that are difficult for humans to access. GenAI's decision is often decisive in these specific scenarios.

When the feedback process is activated, the subgroup that contributes the least to the global consensus is first identified for modification, which is marked as C_{adj}. The preference of GenAI is treated as a reference for preference modification, that is, subgroups who are not within the threshold will modify their preferences toward GenAI's preference. The adjustment direction of subgroups' decisions in Scenario I is depicted in Figure 2. The blue points in a black circle represent a subgroup and the red point represents the centre of subgroup. The green circle represents the adjustment centre range.

Considering experts' tendency to insist on their own preferences, the final modification proposal $p_{ij}^{k'}$ will be generated by combining the initial preference p_{ij}^{k} and the modification reference p_{ij}^{AI}:

$$p_{ij}^{k'} = IFWA\left\langle p_{ij}^{k}, p_{ij}^{AI}\right\rangle = \left\langle 1-\left(1-\mu_{ij}^{k}\right)^{\rho_k}\cdot\left(1-\mu_{ij}^{AI}\right)^{1-\rho_k}, \left(\nu_{ij}^{k}\right)^{\rho_k}\cdot\left(\nu_{ij}^{AI}\right)^{1-\rho_k}\right\rangle \tag{13}$$

where $\rho_k \in [0,1]$ is an adjustment parameter that indicates the degree of reservation of e_k to his/her own preference.

Empirical studies have shown that if there is a GenAI's decision, when regular experts adjust decision outcomes, the influence of authoritative experts diminishes (Zheng et al., 2023). Therefore, if GenAI plays the decisive role, we assume that experts within subgroups have equal weight.

3.4.1.2. Scenario II: GenAI as the equal role. In many decision-making environments, GenAI and experts act side-by-side and together form the core of the decision-making process. For example, in medical diagnosis, GenAI can quickly analyze complex data and provide initial judgments, while doctors make diagnostic decisions by integrating the patient's situation and their own experience. These scenarios emphasize the complementary characteristics of GenAI's decision and experts' decisions, which makes the final decision more comprehensive.

In this scenario, the average preference of GenAI and experts is treated as a reference for preference modification, that is, subgroups who are not within the threshold will modify their preferences toward GenAI-experts' average preference. The adjustment direction of subgroups' decisions in Scenario II is depicted in Figure 3.

Considering experts' tendency to insist on their own preferences, the final modification suggestion $p_{ij}^{k'}$ given to the individual expert will be generated by combining the initial preference p_{ij}^{k} and the modification reference p_{ij}^{DAI}:

$$p_{ij}^{k'} = IFWA\left\langle p_{ij}^{k}, p_{ij}^{DAI}\right\rangle = \left\langle 1-\left(1-\mu_{ij}^{k}\right)^{\rho_k}\cdot\left(1-\mu_{ij}^{DAI}\right)^{1-\rho_k}, \left(\nu_{ij}^{k}\right)^{\rho_k}\cdot\left(\nu_{ij}^{DAI}\right)^{1-\rho_k}\right\rangle \tag{14}$$

where p_{ij}^{DAI} denotes the average decision of GenAI-experts.

An expert is influenced by both GenAI and other experts' decisions. A regular expert is less influenced by authoritative experts when adjusting their decisions. In this scenario, the influence of the centrality

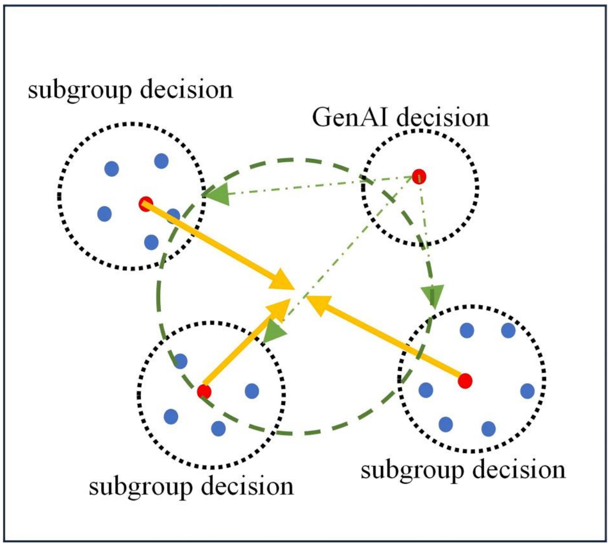


Figure 4. Adjustment direction of subgroups' decisions in Scenario III.

of experts within the same subgroup is reduced, and the expert weights are changed.

The weight of expert e_k is calculated as follows:

$$w_l^k = \frac{\omega_1' \cdot log(1+DT_k) + \omega_2' \cdot log(1+IT_k) + \omega_3' \cdot log(1+CT_k)}{\sum_{h \in c_l} (\omega_1' \cdot log(1+DT_h) + \omega_2' \cdot log(1+IT_h) + \omega_3' \cdot log(1+CT_h))} \tag{15}$$

ω_1', ω_2' and ω_3' corresponds to the adjusted weights of the direct trust value, the indirect trust value, and the centrality value. $\omega_1' + \omega_2' + \omega_3' = 1$, ω_1', ω_2', $\omega_3' > 0$. The log function is used to moderate the magnitude of inconsistency among the direct trust value, the indirect trust value, and the centrality value. In addition, the weight of e_k is normalized to ensure that the sum of weights in each subgroup is 1.

3.4.1.3. Scenario III: GenAI as the auxiliary role. In some decision-making scenarios, GenAI plays the auxiliary role and the final decision remains in hands of human experts. This relationship is particularly notable in domains where experts' experience and intuition are irreplaceable but can benefit from data analysis, such as artistic creation, advanced research and development, and so on (Bell et al., 2024). This process demonstrates the irreplaceability of human experts in complex decisions.

In this scenario, the average preference of experts is treated as a reference for preference modification, that is, subgroups who are not within the threshold will modify their preferences toward the average preference of all subgroups. The adjustment direction of subgroups' decisions in Scenario III is depicted in Figure 4.

Considering experts' tendency to insist on their own preferences, the final modification suggestion $p_{ij}^{k'}$ given to the individual expert will be generated by combining the initial preference p_{ij}^k and the modification reference p_{ij}^{ave}:

$$p_{ij}^{k'} = IFWA\left\langle p_{ij}^k, p_{ij}^{ave} \right\rangle$$
$$= \left\langle 1 - \left(1-\mu_{ij}^k\right)^{\rho_k} \cdot \left(1-\mu_{ij}^{ave}\right)^{1-\rho_k}, \left(\nu_{ij}^k\right)^{\rho_k} \cdot \left(\nu_{ij}^{ave}\right)^{1-\rho_k} \right\rangle \tag{16}$$

Experts refer to GenAI's decision when the consensus is difficult to reach. It has been shown that dissenting participants tend to reflect on their own decisions when it is noted that GenAI's decision and other experts' decisions have consistent results (Zheng et al., 2023). Therefore, in this scenario, we calculate the weight of a subgroup according to the degree of consensus between subgroup's decision and GenAI's decision, and only subgroup consensus is considered.

3.4.2. Feedback model setup

Based on the above analysis, the MCC model is established as follows:

$$Min \sum_{i=1}^{m} \sum_{j=1}^{n} c_k \cdot d(p_{ij}^k, p_{ij}^{k'})$$
$$s.t \begin{cases} GACD \geq \sigma \\ c_k \geq 0 \\ 0 \leq \rho_k \leq 1 \\ 0 \leq \vartheta \leq 1 \\ 0 \leq \sigma \leq 1 \\ e_k \in C_{adj} \end{cases} \tag{17}$$

where σ denote the consensus threshold, and C_{adj} denotes the subgroups that needs to make adjustments. c_k is inversely proportional to the expert's attitude towards GenAI. In this study, $c_k = 1/sa_k$.

3.4.3. Selection process

If the degree of global consensus $GACD$ reaches the threshold σ, the model will go into the selection process.

First, the decision score G_i of x_i is obtained based on the final global preference matrix:

$$G_i = \sum_{j=1}^{n} z_j \cdot p_{ij}^G \tag{18}$$

where p_{ij}^G is the preference value of x_i regarding c_j in the final global preference matrix P^G. In Scenario I and Scenario II, P^G is the average matrix of all subgroups' preferences and GenAI's matrix. In scenario III, P^G is the average matrix of all subgroups' preferences.

Then, according to scoring function of IFN, the final score of x_i is obtained as:

$$S(G_i) = \mu_{G_i} - \nu_{G_i} \tag{19}$$

Finally, all alternatives are ranked according to their scores, and the highest scoring alternative is selected as the best alternative.

3.5. Summarization of the proposed model

The flowchart of our proposed LSGDM model is shown in Figure 5.

Steps of our proposed LSGDM model are as follows.

Step 1: Collect information related to decision making from experts and GenAI.

Step 2: Classify experts into several clusters via our improved DPC algorithm.

Step 3: Determine weights of experts and GenAI by Eqs. (4)–(7).

Step 4: Calculate the degree of consensus: CD_l^k, ACD_l, $AACD_{l-AI}$ and $GACD$.

Step 5: If $GACD \geq \sigma$, then go to selection process; otherwise, select subgroups that need to make modifications and go to next step.

Step 6: Generate heterogeneous feedback adjustment strategies to help experts modify their preferences.

Step 7: Rank alternatives based on global preferences.

4. Case study

Chronic Obstructive Pulmonary Disease (COPD) is one of the most prevalent chronic lung diseases and is currently the fourth leading cause of death. Even experienced medical professionals are difficult to give a definite clinical diagnosis. Fortunately, big data has been used as a valuable source of COPD detection (Rhee, 2021). Previous studies have applied AI to detect COPD features (Wang et al., 2020).

There is an electronic medical record of a COPD patient. The electronic medical record includes basic information about patient and diagnose. To minimize

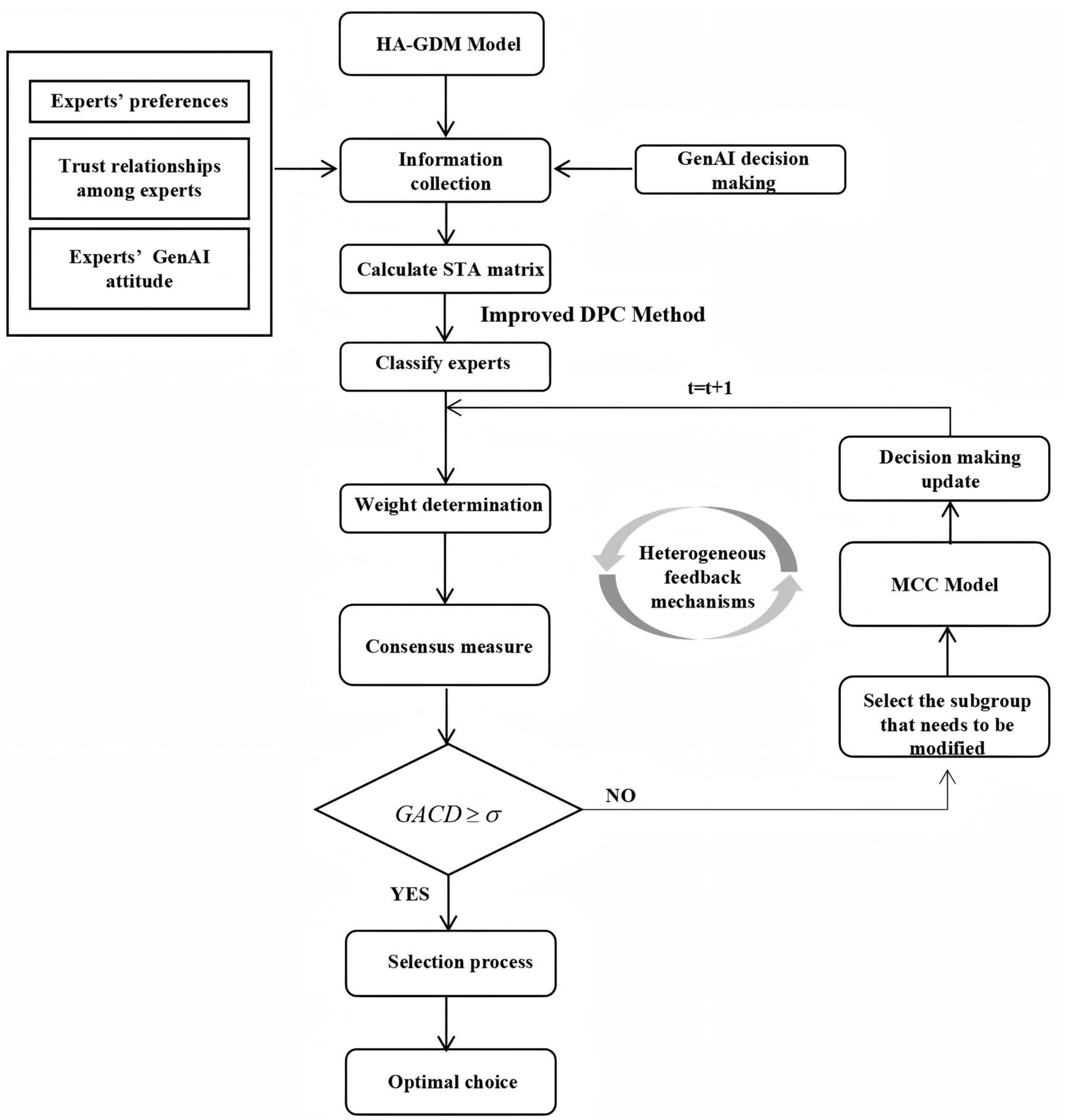


Figure 5. Flowchart of our proposed LSGDM model.

the decision risk, 20 experts $e_k(k = 1, 2, ..., 20)$ from fields of pulmonology, surgery, pharmacology and psychology, respectively, are invited to determine the best alternative for patient's treatment. After careful consultation, four treatment alternatives $X = \{x_1, x_2, x_3, x_4\}$ are identified. x_1 is standard drug therapy including the use of bronchodilators and corticosteroids; x_2 is comprehensive rehabilitation including physiotherapy, nutritional guidance, and psychosocial support; x_3 is a minimally invasive procedure such as lung volume reduction; x_4 is the experimental drug therapy, including the latest development of targeted drugs for COPD. By analyzing these four alternatives, treatment alternatives are given four criteria to be evaluated $C = \{c_1, c_2, c_3, c_4\}$. c_1 is the therapeutic efficacy, which includes symptomatic relief and improvement in lung function; c_2 is the side effect, which includes possible adverse reactions that may occur during course of treatment; c_3 is the cost, which takes into account economic cost of treatment alternative; c_4 is the patient acceptance, which includes patient's preference and compliance with treatment regimen. The weight vector of four criteria is $Z = (0.4, 0.3, 0.2, 0.1)$. GenAI evaluates four treatment alternatives by analyzing a large amount of historical treatment data of COPD patients and then gives a preference score.

To obtain the attitudes of experts toward GenAI, we design a pre-experiment of human-AI coordination, which is shown in Figure 6. A dynamic web application is built to provide core functions for patient diagnostic systems, such as valuation of patient treatment plan outcomes, reasons for plan selection, and reliability of results, based on which an expert can select models and other functions of AI with which he/she can interact.

After appropriate exercises and interactions (it is stipulated that each expert should familiarize himself/herself with the AI system for no less than 30 min and interact with it no less than five times), experts provide their attitudes towards AI by combining the assessment of the accuracy of the AI, the interpretability, and the transparency of the system, and do not change it. Figure 7 displays the overall flowchart.

Figure 8 depicts trust relationships between experts (black lines) and experts' attitude towards GenAI (red numbers). Based on relevant literature and task scenarios studied in this article, parameters are set as follows: d_c=2, α=0.6, η_0=0.5, θ_1=5, σ=0.9, ϑ=0.3, ω_1=0.3, ω_2=0.1, ω_3=0.6, ω'_1=0.6, ω'_2=0.2, ω'_3=0.2.

To find the best alternative, the following steps are carried out.

Step 1. Information collection.

The raw preference matrices of experts and the preference of GenAI are shown in Table 4.

Step 2. Clustering

Using the DPC algorithm, twenty experts are classified into three subgroups. Preferences of subgroups are shown in Table 5.

Step 3. Weight determination

The weights of subgroups are obtained based on Eqs. (4)–(7), as shown in Table 6.

Step 4. Consensus measure

Initial consensus is obtained based on Eqs. (8)–(11), as shown in Table 7.

Step 5. As can be seen from Table 7, $GACD < 0.9$. Thus, the feedback process should be implemented.

Step 6. Activate heterogeneous feedback strategies

(1) Scenario I

According to feedback strategy of Scenario I, experts' preferences are adjusted based on the MCC model. The weights of experts in C_1 are $[1/6,1/6,1/6,1/6,1/6,1/6]^{\mathrm{T}}$; the weights of experts in C_2 are $[0.25,0.25,0.25,0.25]^{\mathrm{T}}$; the weights of experts in C_3 are $[0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1]^{\mathrm{T}}$.

Adjusted consensus is obtained in Table 8.

(2) Scenario II

According to feedback strategy of Scenario II, experts' preferences are adjusted based on the MCC model. The weights of experts in C_1 are $[0.1572,0.1664,0.211,0.1656,0.1468,0.153]^{\mathrm{T}}$; the weights of experts in C_2 are $[0.2526,0.2449,0.2519,0.2506]^{\mathrm{T}}$; the weights of experts in C_3 are $[0.0989,0.0919,0.0993,0.0849,0.1093,0.077,0.1006,0.1072,0.1256,0.1053]^{\mathrm{T}}$.

Adjusted consensus is obtained in Table 9.

(3) Scenario III

According to feedback strategy of Scenario III, experts' preferences are adjusted based on the MCC model. The weights of experts in C_1 are $[0.1647,0.1563,0.1506,0.1692,0.1815,0.1777]^{\mathrm{T}}$; weights of experts in C_2 are $[0.2415,0.2585,0.2479,0.2521]^{\mathrm{T}}$; weights of experts in C_3 are $[0.0927,0.0967,0.1015,0.1068,0.0989,0.0898,0.1122,0.1059,0.1013,0.0942]^{\mathrm{T}}$.

Adjusted consensus is shown in Table 10.

Step 7. Selection process

The final evaluation information for Scenario I is:

$$G_1 :< 0.6311, 0.2162 >,$$
$$G_2 :< 0.5800, 0.2019 >,$$
$$G_3 :< 0.6598, 0.1833 >,$$
$$G_4 :< 0.5780, 0.1885 >.$$

Based on the scoring function of IFN, the following results are obtained: $S(G_1) = 0.4149$, $S(G_2) = 0.3781$, $S(G_3) = 0.4765$, $S(G_4) = 0.3895$.

Therefore, we can obtain the final ranking: $x_3 > x_1 > x_4 > x_2$. The best alternative is x_3.

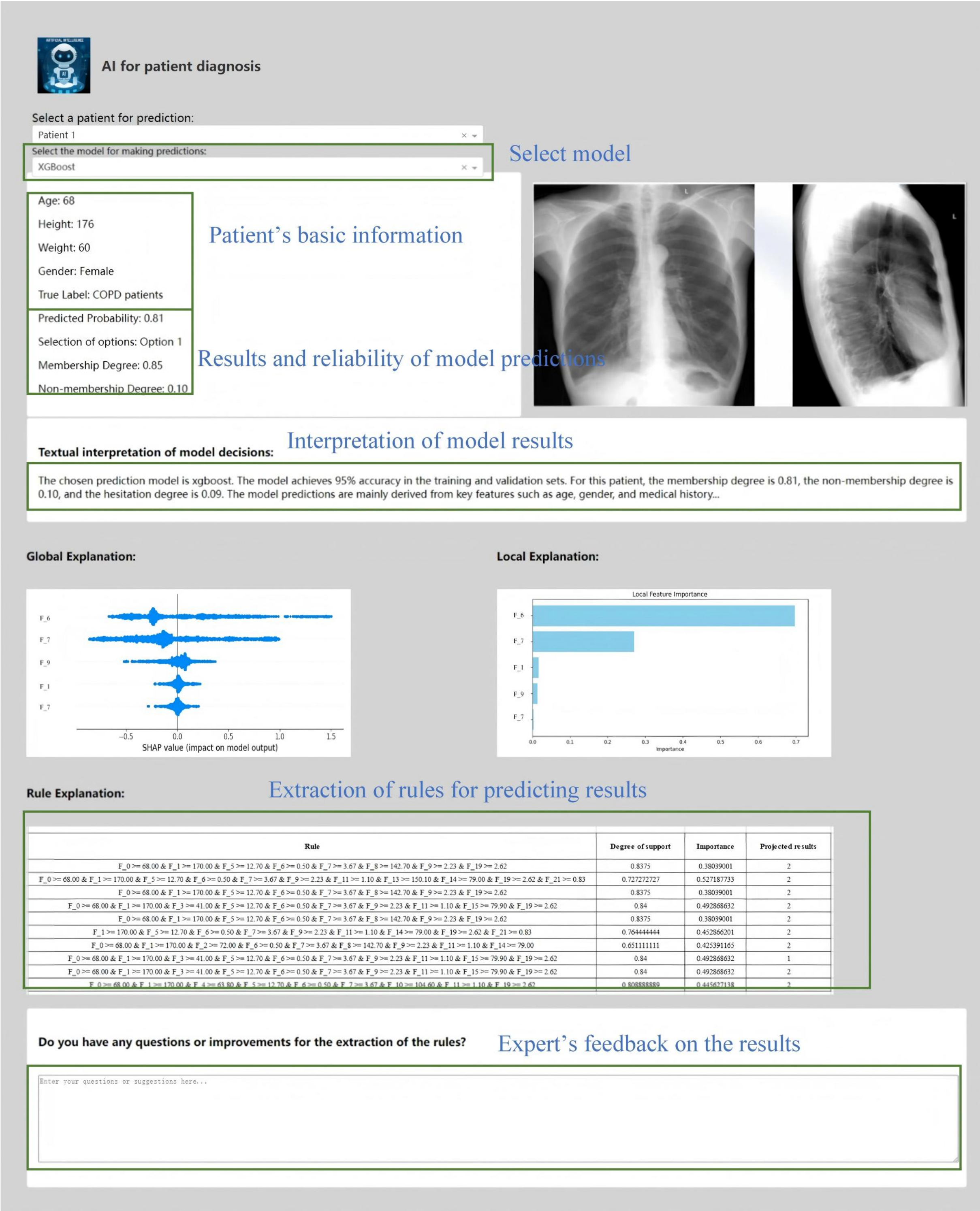


Figure 6. Example of AI decision-making system.

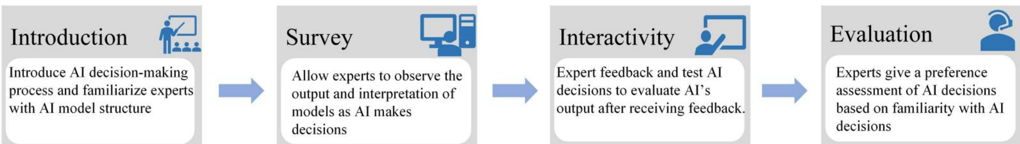


Figure 7. Experimenting over the process of expert adaptation to AI decision making.

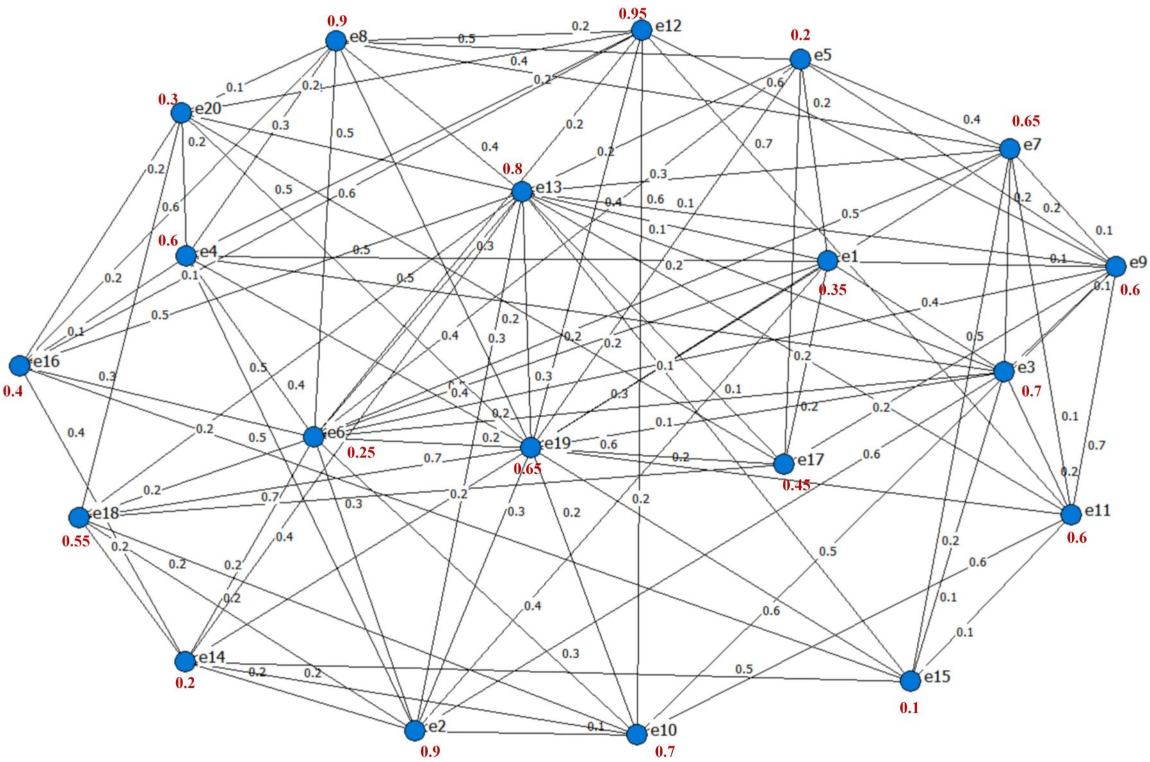


Figure 8. Social trust relationships among experts and experts' attitudes toward GenAI.

Table 4. Preference matrices of experts and GenAI.

	e_1				e_2			
	c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
x_1	<0.4, 0.4>	<0.7,0.2>	<0.8,0.2>	<0.5,0.4>	<0.3, 0.5>	<0.8,0.2>	<0.5, 0.4>	<0.4, 0.1>
x_2	<0.3, 0.4>	<0.4,0.5>	<0.6,0.3>	<0.1,0.7>	<0.3,0.0>	<0.3,0.1>	<0.4,0.0>	<0.9,0.0>
x_3	<0.7,0.2>	<0.3,0.3>	<0.1,0.6>	<0.6,0.2>	<0.4,0.5>	<0.8,0.1>	<1.0,0.0>	<0.2,0.1>
x_4	<0.3,0.3>	<0.5,0.2>	<0.5, 0.1>	<0.8,0.1>	<0.2,0.1>	<0.2,0.1>	<0.4,0.1>	<0.9,0.0>
	e_{20}				e_{AI}			
x_1	<0.7, 0.1>	<0.9, 0.0>	<0.2, 0.7>	<0.6, 0.1>	<0.5, 0.4>	<0.65,0.15>	<0.7,0.15>	<0.7,0.2>
x_2	<0.9,0.0>	<0.6,0.1>	<0.2,0.4>	<0.8,0.0>	<0.65,0.1>	<0.85, 0.1>	<0.4, 0.5>	<0.5, 0.3>
x_3	<1.0,0.0>	<0.4,0.1>	<1.0,0.0>	<0.8,0.1>	<0.55,0.3>	<0.8, 0.2>	<0.8, 0.1>	<0.6, 0.2>
x_4	<0.3,0.0>	<0.3,0.2>	<0.5,0.1>	<0.7,0.3>	<0.9,0.05>	<0.25, 0.6>	<0.65, 0.2>	<0.6, 0.3>

Note: due to the large number of experts, only partial data are presented.

Table 5. Preferences of subgroups.

	p^k			
C_1	<0.5717,0.2837>	<0.4204,0.2415>	<0.6580,0.2966>	<0.8648,0.0609>
	<0.4122,0.2509>	<0.6666,0.1973>	<0.6943,0.1045>	<0.4836,0.2579>
	<0.5745,0.2584>	<0.6690,0.1813>	<0.4402,0.3331>	<0.5489,0.1267>
	<0.4132,0.3001>	<0.5725,0.2183>	<0.3852,0.2921>	<0.5865,0.1431>
C_2	<0.5400,0.3570>	<0.6413,0.2798>	<0.5517,0.3009>	<0.4639,0.2392>
	<0.4112,0.3867>	<0.5939,0.1454>	<0.3863,0.2922>	<0.5911,0.1766>
	<0.5780,0.2485>	<0.7205,0.0924>	<0.5569,0.2007>	<0.6506,0.1338>
	<0.6246,0.1162>	<0.6382,0.1861>	<0.4375,0.2716>	<0.5606,0.2507>
C_3	<0.6369,0.2483>	<0.7501,0.0249>	<0.7826,0.1195>	<0.8501,0.0761>
	<0.5745,0.2584>	<0.6975,0.1758>	<0.6722,0.1987>	<0.5293,0.3206>
	<0.8750,0.0990>	<0.5244,0.2268>	<0.3929,0.3013>	<0.9249,0.0260>
	<0.4727,0.2738>	<0.3509,0.1517>	<0.4211,0.1535>	<0.2503,0.3546>

The final evaluation information for Scenario II is:

$$G_1 :< 0.5907, 0.2657 >,$$

$$G_2 :< 0.5303, 0.2263 >,$$

$$G_3 :< 0.6260, 0.1898 >,$$

$$G_4 :< 0.5688, 0.1926 >.$$

We have $S(G_1) = 0.325$, $S(G_2) = 0.304$, $S(G_3) = 0.4362$, $S(G_4) = 0.3762$. Therefore, $x_3 > x_4 > x_1 > x_2$. The best alternative is x_3.

The final evaluation information for Scenario III is:

$$G_1 :< 0.5655, 0.2903 >,$$

$$G_2 :< 0.5027, 0.2504 >,$$

$$G_3 :< 0.5901, 0.2013 >,$$

$$G_4 :< 0.5440, 0.2074 >$$

We have $S(G_1) = 0.2752$, $S(G_2) = 0.2523$, $S(G_3) = 0.3888$, $S(G_4) = 0.3366$. Therefore, $x_3 > x_4 > x_1 > x_2$. The best alternative is x_3.

Table 6. Subgroup weights.

Subgroup	Number	e_k	Initial weight of subgroup	Initial weight of expert
C_1	6	$e_1, e_5, e_6, e_7, e_{14}, e_{16}$	$w_1 = 0.3226$	$w_1^1 = 0.1521$, $w_2^5 = 0.1572$, $w_2^6 = 0.2525$, $w_2^7 = 0.1616$, $w_1^{14} = 0.1348$, $w_1^{16} = 0.1418$
C_2	4	$e_9, e_{10}, e_{11}, e_{12}$	$w_2 = 0.2976$	$w_3^9 = 0.2597$, $w_3^{10} = 0.239$, $w_3^{11} = 0.2493$, $w_1^{12} = 0.252$
C_3	10	$e_2, e_3, e_4, e_8, e_{13}, e_{15}, e_{17}, e_{18}, e_{19}, e_{20}$	$w_3 = 0.3798$	$w_1^2 = 0.0945$, $w_1^3 = 0.0851$, $w_1^4 = 0.0955$, $w_3^8 = 0.0757$, $w_1^{13} = 0.1246$, $w_1^{15} = 0.0617$, $w_1^{17} = 0.0944$, $w_1^{18} = 0.1074$, $w_1^{19} = 0.158$, $w_2^{20} = 0.1031$

Table 7. Initial consensus.

Global consensus	Consensus between subgroup and GenAI	Consensus of subgroup	Consensus of expert
$GACD = \min\{AACD_{l-AI}, ACD_l\} = 0.7435$	$AACD_{1-AI} = 0.7631$	$ACD_1 = 0.9058$	$CD_1^1 = 0.7828$, $CD_2^5 = 0.749$, $CD_2^6 = 0.7346$, $CD_2^7 = 0.7877$, $CD_1^{14} = 0.7766$, $CD_1^{16} = 0.7656$
	$AACD_{2-AI} = 0.8061$	$ACD_2 = 0.892$	$CD_3^9 = 0.7192$, $CD_3^{10} = 0.7252$, $CD_3^{11} = 0.7116$, $CD_1^{12} = 0.6647$
	$AACD_{3-AI} = 0.7435$	$ACD_3 = 0.8916$	$CD_1^2 = 0.606$, $CD_1^3 = 0.6845$, $CD_1^4 = 0.6261$, $CD_3^8 = 0.6963$, $CD_1^{13} = 0.6923$, $CD_1^{15} = 0.6365$, $CD_1^{17} = 0.7934$, $CD_1^{18} = 0.6346$, $CD_1^{19} = 0.6989$, $CD_2^{20} = 0.7329$

Table 8. Adjusted consensus.

Global consensus	Consensus between subgroup and GenAI	Consensus of subgroup	Consensus of expert
$GACD = \min\{AACD_{l-AI}, ACD_l\} = 0.9$	$AACD_{1-AI} = 0.917$	$ACD_1 = 0.938$	$CD_1^1 = 0.8531$, $CD_2^5 = 0.8419$, $CD_2^6 = 0.792$, $CD_2^7 = 0.8367$, $CD_1^{14} = 0.8333$, $CD_1^{16} = 0.8341$
	$AACD_{2-AI} = 0.9$	$ACD_2 = 0.9336$	$CD_3^9 = 0.8942$, $CD_3^{10} = 0.9075$, $CD_3^{11} = 0.8975$, $CD_1^{12} = 0.8827$
	$AACD_{3-AI} = 0.9418$	$ACD_3 = 0.9392$	$CD_1^2 = 0.7794$, $CD_1^3 = 0.8287$, $CD_1^4 = 0.8743$, $CD_3^8 = 0.832$, $CD_1^{13} = 0.8161$, $CD_1^{15} = 0.7618$, $CD_1^{17} = 0.8546$, $CD_1^{18} = 0.8329$, $CD_1^{19} = 0.8399$, $CD_2^{20} = 0.7738$

Table 9. Adjusted consensus.

Global consensus	Consensus between subgroup and GenAI	Consensus of subgroup	Consensus of expert
$GACD = \min\{AACD_{l-AI}, ACD_l\} = 0.9124$	$AACD_{1-AI} = 0.9124$	$ACD_1 = 0.9378$	$CD_1^1 = 0.8456$, $CD_2^5 = 0.8388$, $CD_2^6 = 0.8035$, $CD_2^7 = 0.8352$, $CD_1^{14} = 0.823$, $CD_1^{16} = 0.8254$
	$AACD_{2-AI} = 0.9161$	$ACD_2 = 0.952$	$CD_3^9 = 0.7192$, $CD_3^{10} = 0.7252$, $CD_3^{11} = 0.7116$, $CD_1^{12} = 0.6647$
	$AACD_{3-AI} = 0.9294$	$ACD_3 = 0.9698$	$CD_1^2 = 0.7568$, $CD_1^3 = 0.8185$, $CD_1^4 = 0.8623$, $CD_3^8 = 0.8409$, $CD_1^{13} = 0.8148$, $CD_1^{15} = 0.7238$, $CD_1^{17} = 0.8533$, $CD_1^{18} = 0.8066$, $CD_1^{19} = 0.8344$, $CD_2^{20} = 0.7603$

Table 10. Adjusted consensus.

Global consensus	Consensus between subgroup and GenAI	Consensus of subgroup	Consensus of expert
$GACD = \min\{ACD_l\} = 0.9058$	$AACD_{1-AI} = 0.8093$	$ACD_1 = 0.9058$	$CD_1^1 = 0.7828$, $CD_2^5 = 0.749$, $CD_2^6 = 0.7346$, $CD_2^7 = 0.7877$, $CD_1^{14} = 0.7766$, $CD_1^{16} = 0.7656$
	$AACD_{2-AI} = 0.8507$	$ACD_2 = 0.9516$	$CD_3^9 = 0.7192$, $CD_3^{10} = 0.7252$, $CD_3^{11} = 0.7116$, $CD_1^{12} = 0.6647$
	$AACD_{3-AI} = 0.8474$	$ACD_3 = 0.9571$	$CD_1^2 = 0.6951$, $CD_1^3 = 0.7754$, $CD_1^4 = 0.7707$, $CD_3^8 = 0.7933$, $CD_1^{13} = 0.735$, $CD_1^{15} = 0.6884$, $CD_1^{17} = 0.7898$, $CD_1^{18} = 0.7738$, $CD_1^{19} = 0.7309$, $CD_2^{20} = 0.6847$

Table 11. The degree of consensus without GenAI.

Global consensus	Consensus of subgroup	Consensus of expert
$GACD = \min\{ACD_l\}$ $=0.9089$	$ACD_1=0.9089$	$CD_1^1=0.7828$, $CD_2^5=0.749$, $CD_2^6=0.7346$, $CD_2^7=0.7877$, $CD_1^{14}=0.7766$, $CD_1^{16}=0.7656$
	$ACD_2=0.949$	$CD_3^9=0.7192$, $CD_3^{10}=0.7252$, $CD_3^{11}=0.7116$, $CD_1^{12}=0.6647$
	$ACD_3=0.9525$	$CD_1^2=0.6866$, $CD_1^3=0.774$, $CD_1^4=0.8253$, $CD_3^8=0.7879$, $CD_1^{13}=0.7384$, $CD_1^{15}=0.6858$, $CD_1^{17}=0.7965$, $CD_1^{18}=0.7854$, $CD_1^{19}=0.7489$, $CD_2^{20}=0.6931$

Table 12. Decision-making results in four different scenarios.

Percentage of GenAI's decision	Adjustment cost	Number of subgroup iterations	Ranking order
Scenarios I	16.78	3	$x_3 > x_1 > x_4 > x_2$
Scenarios II	12.43	3	$x_3 > x_4 > x_1 > x_2$
Scenarios III	1.74	1	$x_3 > x_4 > x_1 > x_2$
No-GenAI	\	1	$x_3 > x_4 > x_1 > x_2$

Table 13. Key metrics of MCC models under different thresholds.

Consensus threshold	0.85	0.86	0.87	0.88	0.89	0.9	0.91	0.92
Number of iterations	$2+2$	$2+2$	$2+3$	$2+3$	$2+3$	$3+4$	$3+4$	$4+4$
Adjustment cost	0.44	1.01	2.05	7.18	8.42	12.43	16.77	25.72
Global consensus	0.859	0.86	0.876	0.8877	0.8924	0.9124	0.9144	0.9276

5. Simulations and comparisons

5.1. The impact of GenAI's role

To illustrate the impact of GenAI's role, this section only considers experts without GenAI's involvement in the proposed model. Other settings remain unchanged. The degree of consensus is shown in Table 11.

The final evaluation information is:

$$G_1 :< 0.5692, 0.2875 >,$$
$$G_2 :< 0.5077, 0.2470 >,$$
$$G_3 :< 0.5926, 0.201 >,$$
$$G_4 :< 0.5443, 0.2076 >$$

We have $S(G_1)=0.2817$, $S(G_2)=0.2607$, $S(G_3)=0.3916$, $S(G_4)=0.3367$. Therefore, $x_3 > x_4 > x_1 > x_2$. The best alternative is x_3.

The decision-making results in four different scenarios are given in Table 12.

In terms of the adjustment cost, Scenario I and Scenario II have a higher cost than other scenarios. Moreover, Scenario III and No-GenAI has the smallest subgroup iterations. This is due to the emergence of GenAI's decision as a new reference expert, which has increased discussions among expert groups. Regarding the ranking of alternatives, it can be seen that the order of (x_1, x_4) changes with the level of participation of GenAI. As the participation level of GenAI increases, the gap between x_1 and x_4 decreases (x_1 is gradually close to x_4). Therefore, it can be concluded that GenAI has a certain impact on the decision-making result, which can be used as a reference for experts.

5.2. Impact of different consensus thresholds

In the LSGDM problem, due to the significant differences in experts' backgrounds, knowledge, and professional skills, it is necessary to set the threshold reasonably based on the specificity of task decision-making. In current research, the values for predefined thresholds are usually set to 0.9 (Gong et al., 2024), 0.85 (Gou et al., 2021), and 0.8 (Tang et al., 2020). For some crucial and significant issues, the minimum consensus level should be very high. However, in other extreme cases, such as emergency problems, the degree of consensus reached due to time constraints should be lower (Xu et al., 2015). Since medical decision-making is a high-risk decision, in this study, we set the consensus threshold to 0.9, which is a relatively high value.

This section explores the impact of consensus thresholds. Eight consensus thresholds {0.85, 0.86, …, 0.92} are assigned to σ to explore the performance of the proposed model. Table 13 provides the performance of the proposed model in terms of three key metrics: number of iterations, adjustment cost and the degree of final global consensus. In addition, to better reflect the efficiency of the CRP, the number of GenAI iterations is also provided. For example, "$1+2$" indicates one iteration of subgroups and two iterations of GenAI.

As can be seen from Table 13, the number of iterations gradually increases as the consensus threshold grows. This means that the higher the requirement for global consensus, the more modifications need to be made. However, it is not difficult to find that no matter the level of the consensus threshold, the model can effectively improve global consensus within a limited number of iterations. Even if σ=0.92, only ($4+4$) iterations are needed to fulfil consensus requirement. These results show that our proposed model can effectively help experts reach consensus and thus solve the LSGDM problem.

5.3. Impact of the unit adjustment cost

Scholars have discussed the impact of the unit adjustment cost on decision results. For instance, Wu et al. (2022) found that the adjustment cost affects initial clustering result and the group consensus's achievement level. However, the current research based on the unit adjustment cost is mostly obtained based on subjective settings. This study relies on experts' attitude towards GenAI when determining the unit adjustment cost, which makes it more in line with real-world human-AI interactive setting. Taking C_2 as an example, Table 14 describes the impact of different unit adjustment costs under Scenario II.

Experts' attitude towards GenAI in C_2 is $\{0.6, 0.7, 0.6, 0.95\}$. The first column of Table 14 provides different cost functions associated with SA. It can be seen that adjustment costs under different functions yield different adjustment results. Compared to the unified cost method, the methods based on four functions all substantially increase the amount of adjustment for e_{12} (the third expert in the bracket, whose decision attitude towards GenAI is relatively optimistic), and moderately decrease the amount of adjustment for e_9, e_{10} and e_{11}. This result is consistent with the fact that experts with AI aversion have a weaker willingness to modify in the feedback process. The total adjustment costs for the first iteration are 3.51, 3.49, 3.51, 3.52, and 3.68, respectively, under each method for the unit cost of $\{1.67, 1.43, 1.67, 1.05\}$. It can be seen that the method for different unit adjustment costs reduces the total costs, which verifies the validity of using the heterogeneous adjustment feedback strategies.

5.4. Comparisons with existing LSGDM models

To verify the superiority of the proposed model, a comparative analysis with existing LSGDM models is carried out. The results are shown in Table 15. The LSGDM model in this study takes the AI perspective into account, making it a unique feature. We first explore how to incorporate different types of interactions between human experts and AI in the CRP and propose corresponding feedback strategies. Therefore, the model proposed in this study is in line with the mainstream trend of AI's decision being increasingly important. Furthermore, our model sets heterogeneous unit adjustment costs for different experts according to experts' attitude towards AI, which also improves upon the previous literature of subjectively setting unit adjustment costs. Our model also has a comprehensive perspective when clustering experts. Most of the studies did not consider the effect of experts' social relations, such as the social status and the degree of trust.

6. Discussions and conclusions

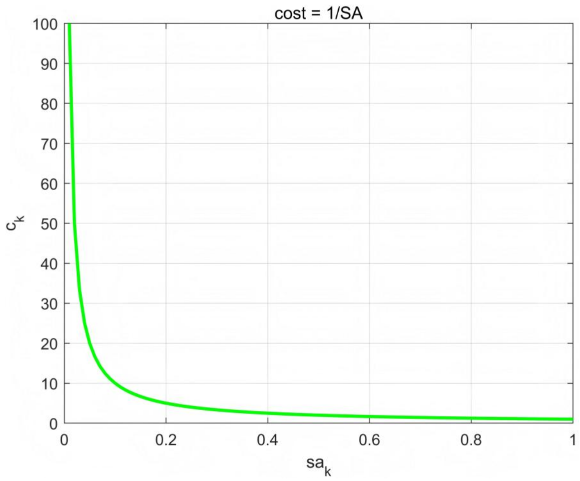
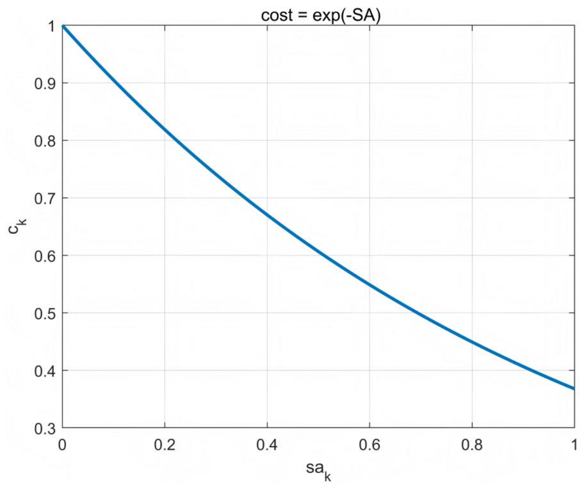
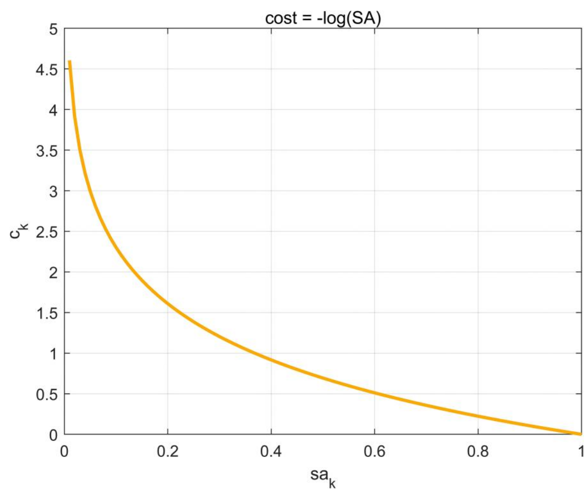
We proposed a new LSGDM model including the participation of GenAI, and analyzed CRP of LSGDM when GenAI has different degrees of participation in decision making. First, the model used an improved DPC algorithm to classify experts. Second, initial weights of experts were determined based on the internal influence of experts, and the weights of subgroups and GenAI are calculated based on intuitionistic fuzzy entropy. In addition, the consensus at different levels is calculated, and a MCC model for heterogeneous feedback strategies in different scenarios is proposed. To validate the feasibility and effectiveness of the model, a medical diagnosis example with comparative analysis is performed. We found that GenAI has different impacts under different levels of GenAI's participation, and heterogeneous adjustment strategies can effectively reduce the adjustment cost.

Our proposed model is highly adaptable. While we use a medical case study for validation, the model is equally applicable to other domains. For instance, in judicial decision making, AI can be trained to provide legal recommendations by analyzing relevant laws and judicial precedents. AI model can take into account factors such as similar precedent cases, case characteristics, and legal provisions, and AI's decisions can then be compared with the rulings of relevant judges to gather consensus feedback. Additionally, the model can be applied to emergency response plan selection, paper ranking in scientific research competitions, and the evaluation of innovations in enterprise technologies. Depending on the complexity of the decision-making task, different human-AI coordination strategies can be employed.

6.1. Management implications

This study has theoretical and managerial implications. First, the LSGDM model in this paper demonstrates the potential of human-AI coordination. With the rapid development of advanced technique, enterprises and policy makers increasingly rely on data and AI to make decisions (Xu et al., 2024). The participation of GenAI can make up for insufficient cognitive capacity or information overload brought by the decision making of a single party (experts) and improve the ability to solve complex decision-making problems involving various elements. In addition, when dealing with decisions involving multiple stakeholders, adding GenAI's decision can improve fairness and diversity of decisions. Second, the feedback strategy plays a significant role in coordinating GenAI decision making with expert decision making. Until now, human-AI coordination has been performed in various fields (De Véricourt

Table 14. Examples of modifications under different adjusted cost scenarios.

Cost functions	Function images	Unit adjustment cost	Adjustment parameter
$c_k = 1/sa_k$		$\{1.67, 1.43, 1.67, 1.05\}$	$\{0.54, 0.61, 0.47, 0.91\}$
$c_k = exp(-sa_k)$		$\{0.55, 0.50, 0.55, 0.37\}$	$\{0.60, 0.54, 0.43, 0.95\}$
$c_k = -log(sa_k)$		$\{0.51, 0.36, 0.51, 0.05\}$	$\{0.61, 0.56, 0.42, 0.94\}$

(*continued*)

et al., 2023; Ge et al., 2021; Zheng et al., 2023), but the decision feedback brought by interpretability and interactivity characteristics of GenAI has not been considered. This study inspires scholars to explore human-AI decision-making outcomes from an interactive feedback perspective. The results indicate that the proposed human-AI coordination model based on heterogeneous feedback strategies can effectively reduce adjustment costs. Meanwhile, for managers, this means that they can optimize decision-making process by adjusting decision feedback strategy, thereby more effectively utilizing resources and time.

The limitation of ethical issues regarding AI has always been a central focus of scholarly research, particularly in high-stake fields such as healthcare and criminal justice (Barrera Ferro et al., 2025). While ethical dilemmas do exist in the use of AI, it

Table 14. Continued.

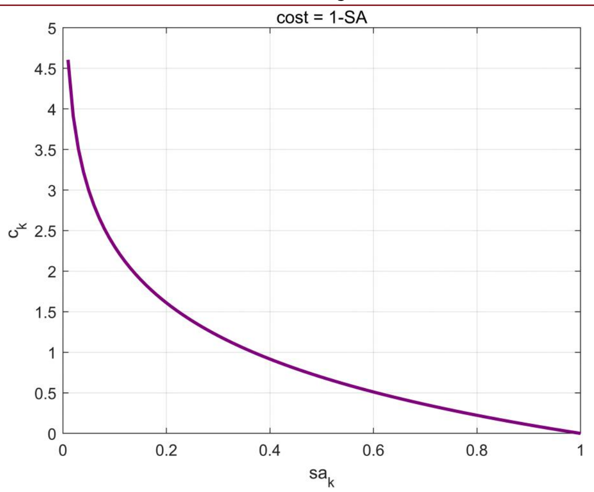
Cost functions	Function images	Unit adjustment cost	Adjustment parameter
$c_k = 1 - sa_k$		$\{0.4, 0.7, 0.4, 0.05\}$	$\{0.58, 0.50, 0.52, 0.92\}$
Unified cost method		$\{1, 1, 1, 1\}$	$\{0.59, 0.65, 0.56, 0.79\}$

Table 15. Comparison with existing LSGDM methods.

Methods	AI decision	The basis of clustering	Clustering method	Subgroup weight-determination method	Minimum cost consensus	Heterogeneous unit adjustment cost
(Tang et al., 2020)	No	Reciprocal preference relation	Fuzzy c-means clustering	Size and cohesion	No	No
(Chen et al., 2022)	No	Experts' preferences	IFN clustering	Consensus value	No	No
(Li et al., 2023)	No	Trust relationship	Spectral clustering	Global weights of DMs	Yes	No
(Liang et al., 2023)	No	Opinion similarity, connectivity similarity and behaviour similarity	community detection based on fuzzy clustering	Global weights of DMs	Yes	Yes
(Gong et al., 2024)	No	Experts' influence	Hierarchical clustering	Consistency degree, subgroup cohesion and subgroup credibility	No	No
(Liang et al., 2024)	No	Opinion similarity and trust relationship	New clustering method	the number of DMs and the combined cohesion of subgroups	No	No
Our method	Yes	Similarity-Trust-Attitude score	Improved density peak clustering	Intuitionistic fuzzy entropy	Yes	Yes

is likely that AI will either complement, coexist with, or even replace current systems, ushering in a new era of AI in healthcare or other fields. In fact, it may be considered unscientific not to use AI (Naik et al., 2022). In response to the current shortcomings in human-AI coordination, the following measures can be taken to effectively reduce the limitation of AI: 1) using diverse and representative data; 2) developing various privacy protection techniques (federal learning, blockchain, etc.); 3) improving the interpretability and transparency of AI systems; 4) improving collaborative communication and education in the use of AI among AI designers and developers, policymakers, and experts. Therefore, prior to implementing the human-AI coordination model proposed in this study, human-centred design principles should be integrated into the AI development process to ensure systematic review and regulatory oversight. Additionally, training and education on AI usage, along with pre-experimental operations, should be conducted to reasonably assess AI's role in decision-making (Schwalbe & Wahl, 2020).

6.2. Limitations and future work

This study still has limitations. First, this study does not address extreme cases where there is a conflict between experts' decisions and GenAI's decision, leaving room for future research into the behavioural dynamics and feedback processes of human-AI decision making within the CRP. Second, the current study assumes that experts' attitudes towards GenAI remain static, potential changes in experts' behaviour influenced by reliance on AI over time can be considered in the future. Third, the current data is small, and future efforts will be made to obtain large hospital-specific data through conducting specific interviews with doctors regarding their willingness to engage with AI in decision making. Fourth, ethical concerns raised by AI have not yet been explored in depth, such as data privacy and the potential for AI bias to influence decision outcomes. Future research will further examine how these factors influence the CRP in LSGDM's human-AI coordination.

Author contribution

Jing Zhang: Conceptualization, Methodology, Software, Writing-original draft, Writing-review and editing. Ning Wang: Funding acquisition, Conceptualization, Supervision, Writing–review and editing. Ming Tang: Methodology, Funding acquisition, Conceptualization, Supervision, Writing–review and editing.

Disclosure statement

No potential competing interest was reported by the authors.

Funding

The work was supported by the National Natural Science Foundation of China. (72401052, 72434001), the National Key R&D Program of China (No. 2021YFC3300200), the Social Science Planning Fund Project of Liaoning Province (No. L23AGL011) and the Fundamental Research Funds for the Central Universities (DUT23RC(3)052).

References

Barrera Ferro, D., Brailsford, S., & Chapman, A. (2025). Improving fairness in machine learning-enabled affirmative actions: A case study in outreach activities in healthcare. *Journal of the Operational Research Society*, *76*(2), 363–374. https://doi.org/10.1080/01605682.2024.2354364

Beilby, K., & Hammarberg, K. (2024). ChatGPT: A reliable fertility decision-making tool? *Human Reproduction (Oxford, England)*, *39*(3), 443–447. https://doi.org/10.1093/humrep/dead272

Bell, J. J., Pescher, C., Tellis, G. J., & Füller, J. (2024). Can AI help in ideation? A theory-based model for idea screening in crowdsourcing contests. *Marketing Science*, *43*(1), 54–72. https://doi.org/10.1287/mksc.2023.1434

Ben-Arieh, D., & Easton, T. (2007). Multi-criteria group consensus under linear cost opinion elasticity. *Decision Support Systems*, *43*(3), 713–721. https://doi.org/10.1016/j.dss.2006.11.009

Boyacı, T., Canyakmaz, C., & de Véricourt, F. (2024). Human and machine: The impact of machine input on decision making under cognitive limitations. *Management Science*, *70*(2), 1258–1275. https://doi.org/10.1287/mnsc.2023.4744

Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239. https://doi.org/10.1002/bdm.2155

Chen, X., Zhang, W., Xu, X., & Cao, W. (2022). A public and large-scale expert information fusion method and its application: Mining public opinion via sentiment analysis and measuring public dynamic reliability. *Information Fusion*, *78*, 71–85. https://doi.org/10.1016/j.inffus.2021.09.015

Choudhary, V., Marchetti, A., Shrestha, Y. R., & Puranam, P. (2025). Human-AI ensembles: When can they work? *Journal of Management*, *51*(2), 536–569. https://doi.org/10.1177/01492063231194968

De Véricourt, F., Gurkan, H., De, V'., & Ericourt, F. (2023). Is your machine better than you? You may never know. *Management Science*. https://doi.org/10.1287/mnsc.2023.4791

Du, Z., Yu, S., & Cai, C. (2024). Constrained community detection and multistage multicost consensus in social network large-scale decision-making. *IEEE Transactions on Computational Social Systems*, *11*(1), 997–1012. https://doi.org/10.1109/TCSS.2023.3265701

Garcia, D., Tolvanen, J., & Wagner, A. K. (2024). Strategic responses to algorithmic recommendations: Evidence from hotel pricing. *Management Science*. https://doi.org/10.1287/mnsc.2022.03740

Ge, R., Zheng, Z., Tian, X., & Liao, L. (2021). Human–robot interaction: When investors adjust the usage of robo-advisors in peer-to-peer lending. *Information Systems Research*, *32*(3), 774–785. https://doi.org/10.1287/isre.2021.1009

Gong, K., Ma, W., Lei, W., & Goh, M. (2024). A dynamic trust network and influence measure-based consensus model for large-scale group decision-making with incomplete intuitionistic fuzzy preference relations. *Journal of the Operational Research Society*, *75*(6), 1157–1177. https://doi.org/10.1080/01605682.2023.2237987

Gou, X., Xu, Z., Liao, H., & Herrera, F. (2021). Consensus model handling minority opinions and noncooperative behaviours in large-scale group decision-making under double hierarchy linguistic preference relations. *IEEE Transactions on Cybernetics*, *51*(1), 283–296. https://doi.org/10.1109/TCYB.2020.2985069

Hermann, E., & Puntoni, S. (2024). Artificial intelligence and consumer behavior: From predictive to generative AI. *Journal of Business Research*, *180*, 114720. https://doi.org/10.1016/j.jbusres.2024.114720

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Information Systems Research*, *32*(3), 713–735. https://doi.org/10.1287/isre.2020.0980

Laker, L. F., Froehle, C. M., Windeler, J. B., & Lindsell, C. J. (2018). Quality and efficiency of the clinical decision-making process: Information overload and emphasis framing. *Production and Operations*

Management, *27*(12), 2213–2225. https://doi.org/10.1111/poms.12777

Lareyre, F., Yeung, K. K., Guzzi, L., Di Lorenzo, G., Chaudhuri, A., Behrendt, C. A., Spanos, K., & Raffort, J. (2023). Artificial intelligence in vascular surgical decision making. *Seminars in Vascular Surgery*, *36*(3), 448–453. https://doi.org/10.1053/j.semvascsurg.2023.05.004

Liang, X., Guo, J., & Liu, P. (2024). A consensus model considers managing manipulative and overconfident behaviours in large-scale group decision-making. *Information Sciences*, *654*, 119848. https://doi.org/10.1016/j.ins.2023.119848

Liang, Y., Ju, Y., Qin, J., Pedrycz, W., & Dong, P. (2023). Minimum cost consensus model with loss aversion based large-scale group decision making. *Journal of the Operational Research Society*, *74*(7), 1712–1729. https://doi.org/10.1080/01605682.2022.2110002

Liang, H., Kou, G., Dong, Y., Chiclana, F., & Herrera-Viedma, E. (2022). Consensus reaching with minimum cost of informed individuals and time constraints in large-scale group decision-making. *IEEE Transactions on Fuzzy Systems*, *30*(11), 4991–5004. https://doi.org/10.1109/TFUZZ.2022.3165373

Li, L., Jiao, S., Shen, Y., Liu, B., Pedrycz, W., Chen, Y., & Tang, X. (2023). A two-stage consensus model for large-scale group decision-making considering dynamic social networks. *Information Fusion*, *100*, 101972. https://doi.org/10.1016/j.inffus.2023.101972

Li, S., Peng, Y., Shen, Y., Zhao, S., Shao, H., Bin, G., Guo, Y., Yang, X., & Fan, C. (2024). Rolling bearing fault diagnosis under data imbalance and variable speed based on adaptive clustering weighted oversampling. *Reliability Engineering & System Safety*, *244*, 109938. https://doi.org/10.1016/j.ress.2024.109938

Li, L., Zhang, H., Peng, H., & Yang, Y. (2018). Nearest neighbors based density peaks approach to intrusion detection. *Chaos, Solitons & Fractals*, *110*, 33–40. https://doi.org/10.1016/j.chaos.2018.03.010

Ma, S., Lei, Y., Wang, X., Zheng, C., Shi, C., Yin, M., & Ma, X. (2023). *Who should I trust: AI or myself? Leveraging human and AI correctness likelihood to promote appropriate trust in AI-assisted decision-making.* Conference on Human Factors in Computing Systems – Proceedings, 1–19. https://doi.org/10.1145/3544548.3581058

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, *175*, 121390. https://doi.org/10.1016/j.techfore.2021.121390

Mahmud, H., Islam, A. K. M. N., Luo, X., & Mikalef, P. (2024). Decoding algorithm appreciation: Unveiling the impact of familiarity with algorithms, tasks, and algorithm performance. *Decision Support Systems*, *179*, 114168. https://doi.org/10.1016/j.dss.2024.114168

Malone, T., Rus, D., & Laubacher, R. (2020). *Artificial intelligence and the future of work.* MIT Work of the Future Research Brief. RB17–2020.

Meng, F. Y., Zhao, D. Y., Gong, Z. W., Chu, J. F., Pedrycz, W., & Yuan, Z. (2024). Consensus adjustment for multi-attribute group decision making based on cross-allocation. *European Journal of Operational Research*, *318*(1), 200–216. https://doi.org/10.1016/j.ejor.2024.05.003

Meng, F., Zhao, D., & Zhang, X. (2023). A fair consensus adjustment mechanism for large-scale group decision making in term of Gini coefficient. *Engineering Applications of Artificial Intelligence*, *126*, 106962. https://doi.org/10.1016/j.engappai.2023.106962

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, *58*(3), 103434. https://doi.org/10.1016/j.im.2021.103434

Naik, N., Hameed, B. M. Z., Shetty, D. K., Swain, D., Shah, M., Paul, R., Aggarwal, K., Brahim, S., Patil, V., Smriti, K., Shetty, S., Rai, B. P., Chlosta, P., & Somani, B. K. (2022). Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility? *Frontiers in Surgery*, *9*, 862322. https://doi.org/10.3389/fsurg.2022.862322

Patel, B. N., Rosenberg, L., Willcox, G., Baltaxe, D., Lyons, M., Irvin, J., Rajpurkar, P., Amrhein, T., Gupta, R., Halabi, S., Langlotz, C., Lo, E., Mammarappallil, J., Mariano, A. J., Riley, G., Seekins, J., Shen, L., Zucker, E., & Lungren, M. (2019). Human–machine partnership with artificial intelligence for chest radiograph diagnosis. *NPJ Digital Medicine*, *2*(1), 111. https://doi.org/10.1038/s41746-019-0189-7

Ray, T. R., Kellogg, R. T., Fargen, K. M., Hui, F., & Vargas, J. (2023). The perils and promises of generative artificial intelligence in neurointerventional surgery. *Journal of Neurointerventional Surgery*, *16*(1), 4–7. https://doi.org/10.1136/jnis-2023-020353

Rhee, C. K. (2021). Chronic obstructive pulmonary disease research by using big data. *The Clinical Respiratory Journal*, *15*(3), 257–263. https://doi.org/10.1111/crj.13305

Saban, M., & Dubovi, I. (2024). A comparative vignette study: Evaluating the potential role of a generative AI model in enhancing clinical decision-making in nursing. *Journal of Advanced Nursing*. https://doi.org/10.1111/jan.16101

Schwalbe, N., & Wahl, B. (2020). Artificial intelligence and the future of global health. *Lancet (London, England)*, *395*(10236), 1579–1586. https://doi.org/10.1016/S0140-6736(20)30226-9

Selby, D., Spriestersbach, K., Iwashita, Y., Bappert, D., Warrier, A., Mukherjee, S., Asim, M. N., Kise, K., & Vollmer, S. (2024). *Quantitative knowledge retrieval from large language models.* arxiv preprint arxiv:2402.07770.

Sorin, V., Klang, E., Sklair-Levy, M., Cohen, I., Zippel, D. B., Balint Lahat, N., Konen, E., & Barash, Y. (2023). Large language model (ChatGPT) as a support tool for breast tumour board. *NPJ Breast Cancer*, *9*(1), 44. https://doi.org/10.1038/s41523-023-00557-8

Sturm, T., Gerlach, J. P., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., & Buxmann, P. (2021). Coordinating human and machine learning for effective organization learning. *MIS Quarterly*, *45*(3), 1581–1602. https://doi.org/10.25300/MISQ/2021/16543

Tang, M., & Liao, H. (2021). From conventional group decision making to large-scale group decision making: What are the challenges and how to meet them in big data era? A state-of-the-art survey. *Omega*, *100*, 102141. https://doi.org/10.1016/j.omega.2019.102141

Tang, M., Liao, H., Xu, J., Streimikiene, D., & Zheng, X. (2020). Adaptive consensus reaching process with hybrid strategies for large-scale group decision making.

European Journal of Operational Research, *282*(3), 957–971. https://doi.org/10.1016/j.ejor.2019.10.006

Te'eni, D., Yahav, I., Zagalsky, A., Schwartz, D., Silverman, G., Cohen, D., Mann, Y., & Lewinsky, D. (2023). Reciprocal human-machine learning: A theory and an instantiation for the case of message classification. *Management Science*. https://doi.org/10.1287/mnsc.2022.03518

Thuy, A., & Benoit, D. F. (2024). Explainability through uncertainty: Trustworthy decision-making with neural networks. *European Journal of Operational Research*, *317*(2), 330–340. https://doi.org/10.1016/j.ejor.2023.09.009

Tu, Y., Song, J., Xie, Y., Zhou, X., & Lev, B. (2024). Facilitating large-scale group decision-making in social networks: A bi-level consensus model with social influence. *Information Fusion*, *105*, 102258. https://doi.org/10.1016/j.inffus.2024.102258

Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., Denniston, A. K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B. A., Mathur, P., McCradden, M. D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D. S. W., … McCulloch, P. (2022). Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature Medicine*, *28*(5), 924–933. https://doi.org/10.1038/s41591-022-01772-9

Wang, C., Chen, X., Du, L., Zhan, Q., Yang, T., & Fang, Z. (2020). Comparison of machine learning algorithms for the identification of acute exacerbations in chronic obstructive pulmonary disease. *Computer Methods and Programs in Biomedicine*, *188*, 105267. https://doi.org/10.1016/j.cmpb.2019.105267

Wang, B., Yuan, T., & Rau, P. L. P. (2024). Effects of explanation strategy and autonomy of explainable AI on human–AI collaborative decision-making. *International Journal of Social Robotics*, *16*(4), 791–810. https://doi.org/10.1007/s12369-024-01132-2

Wu, T., Liu, X., Qin, J., & Herrera, F. (2022). A new clustering algorithm with preference adjustment cost to reduce the cooperation complexity in large-scale group decision making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *52*(8), 5271–5283. https://doi.org/10.1109/TSMC.2021.3120809

Xu, Z. (2007). Some similarity measures of intuitionistic fuzzy sets and their applications to multiple attribute decision making. *Fuzzy Optimization and Decision Making*, *6*(2), 109–121. https://doi.org/10.1007/s10700-007-9004-z

Xu, X. H., Du, Z. J., & Chen, X. H. (2015). Consensus model for multi-criteria large-group emergency decision making considering non-cooperative behaviours and minority opinions. *Decision Support Systems*, *79*, 150–160. https://doi.org/10.1016/j.dss.2015.08.009

Xu, X., Wang, Y., Cheng, T. C. E., & Choi, T. M. (2024). Should live broadcasting platforms adopt artificial intelligence? A sales effort perspective. *European Journal of Operational Research*, *318*(3), 979–999. https://doi.org/10.1016/j.ejor.2024.05.021

Yu, S. M., Du, Z. J., Zhang, X. Y., Luo, H. Y., & Lin, X. D. (2022). Trust cop-k means clustering analysis and minimum-cost consensus model considering voluntary trust loss in social network large-scale decision-making. *IEEE Transactions on Fuzzy Systems*, *30*(7), 2634–2648. https://doi.org/10.1109/TFUZZ.2021.3089745

Zhang, H., Kou, G., & Peng, Y. (2022). Large-scale peer-to-peer loan consensus based on minimum cost consensus. *Journal of the Operational Research Society*, *73*(10), 2326–2337. https://doi.org/10.1080/01605682.2021.1981782

Zheng, C., Wu, Y., Shi, C., Ma, S., Luo, J., & Ma, X. (2023). *Competent but rigid: Identifying the gap in empowering AI to participate equally in group decision-making*. Conference on Human Factors in Computing Systems – Proceedings, 1–19. https://doi.org/10.1145/3544548.3581131