Multi-scale Group Decision-Making Employing Large Language Model for Sentiment-Oriented Grouping

JOSÉ RAMÓN TRILLO[1], MANUEL JESUS COBO[1], IGNACIO JAVIER PÉREZ[1], JUAN ANTONIO MORENTE-MOLINERA[1], FRANCISCO JAVIER CABRERIZO[1], and ENRIQUE HERRERA-VIEDMA[1]

[1]Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence, DaSCI, University of Granada, Granada, 18071, Spain

Corresponding author: Ignacio Javier Pérez (e-mail: ijperez@decsai.ugr.es).

This work has been supported by the grant PID2022-139297OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU. Moreover, it is part of the project C-ING-165-UGR23, co-funded by the Regional Ministry of University, Research and Innovation and by the European Union under the Andalusia ERDF Program 2021-2027.

ABSTRACT In contemporary decision-making scenarios involving multiple experts, the expression of opinions through natural language poses significant challenges for traditional analytical frameworks. As online collaboration expands, so does the heterogeneity of expert evaluations, both in scale and form, particularly when assessments are expressed via unstructured textual comments. Conventional Multi-scale Group Decision-Making methods, while effective in managing diverse evaluation metrics, often fall short in processing the semantic complexity of human language. This study proposes an enhanced Multi-scale Group Decision-Making framework that integrates a Large Language Model to classify experts' comments with greater contextual sensitivity. Unlike traditional sentiment analysis methods based on fixed lexical rules, Large Language Models capture deeper linguistic nuancessuch as tone, intent, and implied meaningallowing for more accurate sentiment-oriented grouping of experts based on their opinions. These refined groups contribute to a more coherent consensus process and lead to improved decision outcomes. The proposed approach advances group decision methodologies by bridging structured decision theory with state-of-the-art natural language understanding.

INDEX TERMS Group Decision-Making, Consensus Reaching Process, Grouping, Multi-scale Method

I. INTRODUCTION

GROUP Decision-Making (GDM) involves the collective evaluation of a finite array of alternatives by a panel of experts, each contributing unique perspectives and domain-specific knowledge [1]–[3]. To systematically manage this diversity of input, a broad spectrum of GDM methodologies has been developed. However, the advent of pervasive digital interconnectivity and the proliferation of online deliberation platforms have profoundly altered the landscape in which these decisions unfold.

This evolution has not only expanded the number of participating experts but has also multiplied the dimensions of the alternatives under scrutiny. Crucially, it has diversified the modalities through which preferences are articulated, shifting from rigid, uniform structures to more flexible and natural modes of expression. In response to this growing complexity, Multi-scale Group Decision-Making (MsGDM) approaches have emerged [4], [5], allowing experts to assess the same alternative using different linguistic or numerical scales. These approaches aim to accommodate the heterogeneity of evaluative criteria and expression styles. However, as the diversity of input continues to increase, traditional MsGDM frameworks are facing growing limitations in terms of integration and interpretability.

In the current literature, however, where decision environments are characterised by heterogeneity, ambiguity, and high-volume textual data, these models require significant refinement. The challenge lies not merely in aggregating numeric evaluations, but in interpreting the semantic richness and implicit reasoning embedded in expert commentary. This requires a methodological shift: the use of advanced computational models that can interpret unstructured language and turn it into actionable insights for decision-making.

Within the MsGDM paradigm, expert deliberation frequently unfolds through discursive contributions articulated in natural language. These commentaries, while rich in nuance and contextual depth, pose inherent challenges to computational interpretation due to their informal structure, semantic ambiguity, and reliance on implicit reasoning [6]. Conventional automated systemspredicated upon formalised, deterministic representations of informationare ill-equipped to navigate the intricacies of such human expression. As a result, much of the latent knowledge embedded in expert discourse is either marginalised or entirely excluded from analytical consideration within traditional MsGDM frameworks.

To mitigate this epistemic loss and more fully capitalise on the informational breadth of expert interaction, the integration of methodologies derived from Natural Language Processing (NLP) becomes imperative [7]. Among these, sentiment analysis constitutes a pivotal instrument, offering a systematic means of discerning the evaluative polarity-positive, negative, or neutralimplicit in textual contributions. Furthermore, sentiment analysis enables the identification of referential targets within commentary, thereby elucidating which specific alternatives are being scrutinised or juxtaposed. This infusion of semantic granularity enhances the interpretive capacity of MsGDM systems, paving the way for more sophisticated, linguistically-informed aggregation of expert perspectives.

Recent research has explored the incorporation of sentiment analysis techniques within MsGDM frameworks [8]. Nevertheless, such approaches frequently depend on rigid classification schemes or sentiment lexicons with limited expressive capacity, thereby constraining their ability to capture the subtlety and depth of expert opinion. In response to these limitations, the present study proposes an innovative methodology that harnesses the capabilities of a Large Language Model (LLM) to conduct a refined sentiment analysis of expert commentary [9].

Through the nuanced linguistic comprehension afforded by the LLM, expert evaluations are semantically analysed and subsequently partitioned into two distinct groups: one comprising individuals whose assessments are predominantly positive in tone, and another characterised by a preponderance of negative sentiment [10]. This strategic segmentation permits more label and context-sensitive evaluations within each subgroup, a critical factor in the pursuit of effective and equitable consensus-building [11].

The integration of LLMs into sentiment analysis represents a methodological advancement over traditional approaches that rely heavily on predefined lexical resources or rule-based classifiers. LLMs are capable of capturing deeper semantic structures, contextual subtleties, and implicit evaluative cues embedded in expert commentary. Their contextual reasoning and generalisation capabilities enable a more nuanced and accurate interpretation of expert sentiment, especially in deliberative settings where opinions are expressed with rhetorical complexity and domain-specific language. Moreover, the employment of an LLM facilitates the dynamic management of lexical and rhetorical variability inherent in expert discourse, thereby enhancing the methodological robustness of the analysis. It also increases procedural efficiency by concentrating deliberative efforts around emergent points of convergence within each group. Upon the intra-group analysis, each groups contribution is weighted using a composite operator that integrates both individual preference intensities and the degree of internal consensus achieved.

This weighted aggregation culminates in a comprehensive and hierarchically structured ranking of alternatives, thereby furnishing a rigorous, data-driven foundation for collective decision-making. Ultimately, the proposed framework augments the management of qualitative information during expert deliberation processes and substantially enhances both the precision and the contextual relevance of the decisions renderedparticularly in scenarios characterized by complexity, uncertainty, and evolving dynamics.

In response to the increasing complexity and ambiguity of expert commentary in group decision-making contexts, this study adopts LLMs to perform refined sentiment analysis. Unlike traditional sentiment analysis approachestypically based on fixed lexicons or rule-based classifiersLLMs provide a more robust and context-aware interpretation of expert opinions by capturing semantic nuance, rhetorical intent, and implicit evaluative signals. This enables the formation of semantically coherent expert groups, which in turn facilitates more accurate preference aggregation and consensus estimation. The main contributions of this paper include: the integration of an LLM-based discourse modelling mechanism for sentiment-oriented expert grouping; the development of a multi-scale aggregation procedure that accounts for intra- and inter-group dynamics; and the validation of the proposed framework through a case study, demonstrating improved interpretability and decision quality in complex environments.

The remainder of this manuscript is structured as follows: Section II lays the theoretical groundwork by exploring the foundational concepts of MsGDM and the incorporation of LLMs as a novel mechanism for enhancing the interpretability of expert discourse. Section III elaborates on the proposed decision-making framework, detailing its procedural architecture and the integration of linguistic intelligence into expert grouping. Section IV presents a case study that exemplifies the application of the model in a real-world decision context. Section V offers a critical appraisal of the systems performance, highlighting its comparative advantages and limitations relative to existing approaches in the literature. Lastly, Section VI synthesises the core contributions of this work and delineates promising avenues for future inquiry.

II. PRELIMINARIES

This section expounds upon the foundational underpinnings of the methodology advanced in this manuscript. It is systematically divided into two subsections: Section II-A elucidates the theoretical constructs and salient principles inherent to Group Decision Making problems. Thereafter, Section II-B offers an erudite exposition of LLMs, delineating their intrinsic architectures and salient functionalities, alongside a discussion of their salient applications that substantiate and enhance the framework proposed herein.

A. GROUP DECISION-MAKING METHODS

Group decision-making is a structured process where a defined set of experts evaluates a finite set of alternatives. Each expert articulates their evaluations and preferences, which are subsequently amalgamated via a rigorous methodological framework aimed at discerning the most suitable alternative(s) under the collective consensus [12]–[14].

To formalize these approaches devised for addressing group decision-making problems, the set of experts is denoted as $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_\Delta\}$, where $\Delta \in \mathbb{N}$ represents the total number of participating experts. Correspondingly, the set of alternatives is represented as $A = \{a_1, a_2, \dots, a_\omega\}$, with $\omega \in \mathbb{N}$ denoting the number of feasible options subject to evaluation [8], [15], [16].

Humanmachine interaction is challenging because humans and machines process information differently. Whereas machines fundamentally operate on numerical data, human communication predominantly transpires through symbolic and natural language systems. In an effort to reconcile this divergence, numerical modelling techniques have been employed to transmute human judgments into a machine-interpretable format [17]–[19]. Nonetheless, this translation often entails considerable complexity. The numerical labels envisioned by an expert may diverge substantially from those predefined within the modelling schema, thereby constraining the expert's capacity to convey nuanced assessments, whether more precise or more generalised, with fidelity.

To surmount this limitation, it becomes imperative to afford experts the autonomy to select their own set of numerical labels when appraising alternatives. Such flexibility ensures that expert input remains both precise and semantically rich. Consequently, the adoption of an MsGDM framework is warranted, enabling the accommodation of varying levels of label in numerical expressions, thus reflecting the heterogeneous modes through which experts elect to articulate their evaluations [20]–[22].

The multi-scale numeric paradigm comprises a series of pivotal phases:

- **Acquisition of Expert Judgments:** Domain specialists articulate their evaluations employing heterogeneous numeric label sets, thereby facilitating a nuanced and precise expression of subjective preferences beyond the constraints of a monolithic scale.
- **Harmonisation of Inputs:** Subsequent to the compilation of expert assessments, a rigorous normalisation procedure is undertaken to transmute the heterogeneous numeric representations into a unified and coherent scale, thus ensuring the integrity and comparability of the aggregated data.
- **Computational Synthesis and Reconfiguration:** The harmonised data serves as the substrate for the ensuing analytical operations. Where the exposition of results demands adherence to an alternative numeric labelling schema, the synthesised information is meticulously reconfigured to align with the prescribed representational framework.

A myriad of multi-scale numeric models have been thoroughly investigated within the scholarly corpus, each meticulously devised to address distinct application exigencies. For instance, the model expounded in [23] integrates probabilistic numeric constructs to adeptly accommodate the inherent uncertainty permeating expert judgments, thereby furnishing a flexible and nuanced framework for representing assessments characterised by varying degrees of confidence. Such probabilistic formulations prove especially efficacious when confronted with non-deterministic or ambiguous opinion landscapes. Complementarily, the work of [24] endeavours to augment both the precision and interpretability of supervised classification methodologies by permitting experts to submit evaluations across varying strata of numeric label.

B. LARGE LANGUAGE MODELS

LLMs are a major advance in natural language processing, capable of understanding and generating text with strong contextual and syntactic awareness. Architected predominantly on deep learning frameworks such as the Transformer [9], LLMs are trained on massive and diverse textual corpora, enabling them to internalise complex linguistic structures, semantic nuances, and pragmatic subtleties [25]–[27]. Recent research has also explored GPT-assisted consensus models and LLM-based multi-agent negotiation systems, where generative models are employed to facilitate dialogue, summarise positions, or propose consensus solutions. These frameworks highlight the potential of LLMs not only for classification tasks but also for active participation in deliberative processes.

The principal strength of LLMs resides in their versatility across a broad spectrum of language tasksincluding but not limited to text generation, machine translation, summarisation, question answering, and sentiment classificationwithout the necessity for task-specific architectures. Through mechanisms such as transfer learning and fine-tuning, these models adapt proficiently to specialised domains, thereby serving as pivotal tools in both academic research and industry applications [28]–[30].

In the sphere of group decision-making, LLMs have demonstrated significant potential by enabling the effective processing of unstructured expert inputs, such as textual comments, evaluations, and deliberations. By converting these qualitative data into structured representations, LLMs facilitate their integration within computational decision frameworks, thereby bridging the longstanding gap between symbolic human language and numeric data processing [31], [32]. This capability not only enhances the accuracy of group consensus mechanisms but also allows for richer semantic understanding of expert judgments.

Notable applications of LLMs in decision-making contexts include the work by [31], which employs LLMs to synthesise expert opinions in complex policy formulation scenarios, and the study by [33], where LLMs are leveraged to support interactive decision-making systems that dynamically adapt to evolving user preferences and contextual information. These investigations underscore the transformative role of LLMs in augmenting decision-making processes through improved interpretability and responsiveness. Furthermore, LLMs facilitate the dynamic generation of domain-specific vocabularies and ontologies, supporting more precise classification and grouping of textual data according to semantic attributes. Such advancements contribute significantly to multi-scale decision-making methodologies by accommodating the inherent complexity and heterogeneity of expert evaluations [34].

III. MSGDM EMPLOYING LLM FOR S-OC

This section delineates the structured methodology underpinning the proposed GDM model. The approach integrates semantic inference capabilities offered by LLMs to process and synthesise expert-generated content, facilitating an advanced deliberative mechanism. The methodological pipeline comprises the following sequential components (See Figure 1):

- **Dialogic Articulation of Expert Perspectives**: Experts participate in moderated deliberations, expressing evaluative positions and argumentative contributions related to the decision context under scrutiny.
- **Normalisation of reciprocal preference relations**: Since the assessments issued by the experts may use different numerical sets, it is essential to unify their scale. At this stage, the opinions are transformed through a normalisation process that ensures a homogeneous and comparable distribution of all assessments, preserving the analytical consistency of the model.
- **Latent grouping via LLM-Driven Discourse Modelling**: The refined discourse is subjected to unsupervised grouping techniques guided by high-dimensional representations derived from LLMs. This enables the detection of underlying thematic affinities and evaluative congruence among expert narratives.
- **Quantification of Group Salience**: Each group is assigned a salience score reflecting its epistemic weight within the broader deliberative corpus, based on internal coherence and its influence on the decision space.
- **Consensus Gradient Estimation across Groups**: Inter-group relationships are examined to ascertain the degree of convergence or cognitive dissonance, revealing macroscopic consensus patterns and potential fault lines in collective reasoning.
- **Synthesis of the Reciprocal Collective Preference Structure**: Expert judgments, filtered and aggregated through the semantic groups, are consolidated into a reciprocal collective preference matrix that encodes the bidirectional support among alternatives.
- **Derivation of the Final Preference Ranking**: The global preference structure is exploited to generate an ordinal ranking of the alternatives, ensuring that the resulting prioritisation reflects both local consensus within groups and overarching inter-group alignment.

The subsequent subsections provide an in-depth exposition of each phase, elaborating on both the theoretical rationale and algorithmic implementation (see Algorithm 1).

The feedback mechanism (lines 1719) detects when inter-group consensus CG falls below the threshold α. In such cases, experts are invited to revise their assessments for up to ρ iterations, progressively narrowing differences until $CG \geq \alpha$. Lines 20-24 then aggregate the updated group matrices into the final collective preference structure and compute the QGDD scores that yield the final ranking. For reproducibility, we provide a minimal parameter configuration that allows readers to replicate the sentiment classification and grouping steps. The following settings were used in our implementation: $model = roberta - base$, $tokenizer = RoBERTaTokenizer$, maximum sequence $length = 128$, classification threshold $(\theta) = 0.80$, consensus threshold $(\alpha) = 0.93$, $batchsize = 32$, learning $rate = 2e - 5$, and training $epochs = 4$. These parameters, together with the pseudocode provided in Algorithm 1, enable the reproduction of the key results reported in the case study.

A. DIALOGIC ARTICULATION OF EXPERT PERSPECTIVES

Once the set of experts and the collection of feasible alternatives have been formally established, the subsequent phase pertains to the specification of the evaluative mechanism through which experts articulate their preferences. Typically, this expression is facilitated via structured representations such as linguistic term sets or numerical scales, both of which enable comparative or absolute judgments. In the present framework, a numerical schema is adopted, thereby allowing for the formulation of reciprocal preference structures that encapsulate the evaluative stance of each expert.

Specifically, the evaluative information elicited from deliberative exchangescomprising the verbal contributions and argumentation of the participating expertsis systematically captured and incorporated into the decision-making apparatus.

The resulting preference data are formalised through reciprocal preference matrices of dimension $\omega \times \omega$, where ω denotes the number of alternatives under consideration. Each matrix, denoted by Ω_k, corresponds to an individual expert $k \in \{1, \ldots, \Delta\}$. The entries s_k^{rt} of Ω_k encode the degree of preference that expert k assigns to alternative a_r over a_t, under the condition $r \neq t$. The main diagonal is left undefined, as self-comparison lacks semantic relevance.

Formally, each preference value is derived through an expert-specific mapping $\mu_k : A \times A \rightarrow [0, I_k]$, where $I_k \in \mathbb{N}$ denotes the cardinality of the experts numeric scale. Thus,

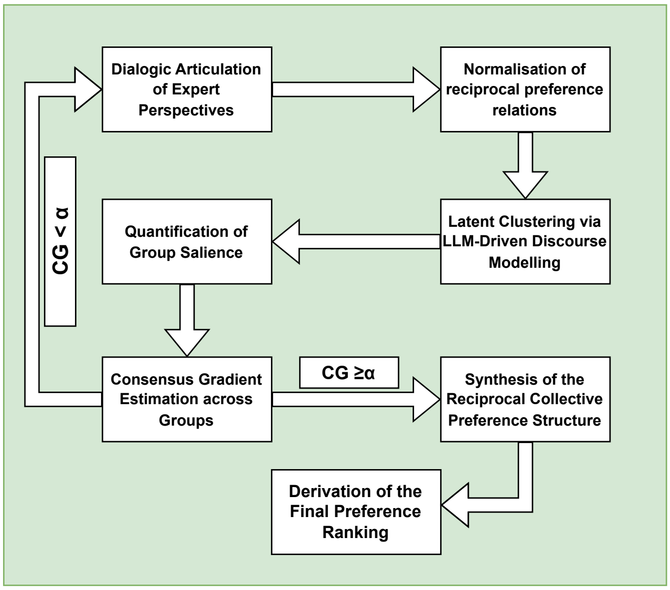

FIGURE 1: Flowchart of the proposed multi-scale group decision-making framework. To complement the figure, Section III provides a detailed breakdown of the framework into six sub-modules, each with its corresponding inputs and outputs.

the matrix is defined as:

$$\Omega_k = (s_k^{rt};\ r \neq t = 1, \ldots, \omega), \quad \text{where} \quad s_k^{rt} = \mu_k(a_r, a_t), \quad \text{and} \quad s_k^{rt} + s_k^{tr} = I_k. \tag{1}$$

This reciprocity condition ensures internal consistency and facilitates subsequent aggregation and analysis of collective preferences.

B. NORMALISATION OF RECIPROCAL PREFERENCE RELATIONS

To establish a coherent framework for comparative evaluation, it is imperative to normalise the preference values s_k^{rt} provided by each expert to a uniform scale. Each element s_k^{rt} embodies the degree of preference expressed by expert t for alternative a_r relative to alternative a_t. The normalisation process involves mapping these heterogeneous values onto a standardised range bounded by zero and the maximum permissible score I_k specific to each expert. The normalized preference S_k^{rt} is formalized as follows: $S_k^{rt} = \frac{s_k^{rt}}{I_k}$. This operation effectively rescales all preference evaluations to the unit interval $[0, 1]$, wherein a value of zero signifies minimal preference and unity denotes maximal endorsement. Such transformation mitigates discrepancies arising from differing evaluation scales, thereby enabling equitable comparison across experts.

The entirety of the normalised preferences is then structured into the matrix $B_k = (S_k^{rt}; \quad r \neq t = 1, \ldots, \omega)$, which encapsulates the pairwise comparative judgments of expert k within a consistent evaluative framework. This matrix serves as a fundamental input for subsequent aggregation and synthesis steps within the group decision-making methodology.

C. LATENT GROUPING VIA LLM-DRIVEN DISCOURSE MODELLING

Let us consider a set of Δ experts, each providing some comments $c_{bi} \in \mathcal{CC}_i$, where $\mathcal{CC}_i = \{c_{1i}, c_{2i}, \ldots, c_{u_i i}\}$ is the compilation of the textual contributions of i-expert made during the group's decision-making process. Our objective is to classify these comments according to their sentiment orientation (positive or negative) and to aggregate their corresponding preference relations accordingly (See Figure 2).

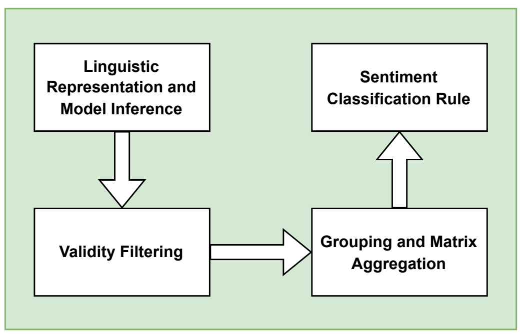

FIGURE 2: Flow chart of latent grouping through an LLM-based discourse model.

The choice of RoBERTa as the backbone model is motivated by its proven effectiveness in sentiment analysis

Algorithm 1 Multi-scale Group Decision-Making with LLM-Based Sentiment-Oriented Grouping

Require: Set of experts $\Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_\Delta\}$; Set of alternatives $A = \{a_1, a_2, \ldots, a_\omega\}$; Expert comments $CC = \{CC_1, CC_2, \ldots, CC_\Delta\}$; Reciprocal preference matrices Ω_k

Ensure: Final ranking of alternatives

for each expert $\gamma_k \in \Gamma$ **do**
  Normalize Ω_k to obtain B_k {Convert each experts preferences to unified [0,1] scale}
end for
for each comment $c \in CC$ **do**
  Tokenize and embed c using a pre-trained LLM (e.g., RoBERTa) {Create contextual embeddings of comments}
  Compute sentiment probabilities via softmax output {Obtain positive/negative sentiment scores}
  Apply sentiment classification rule with threshold θ and keyword filters W_p, W_m {Classify as positive, negative, or uncertain}
end for
Group experts based on sentiment polarity {Separate experts into positive (Ψ_p) and negative (Ψ_m)}
$\Psi_p \leftarrow \{\gamma_k : f(c_k) = \text{pos}\}$
$\Psi_m \leftarrow \{\gamma_k : f(c_k) = \text{neg}\}$
Aggregate normalized matrices {Build group-level preference matrices}
$G_p = \frac{1}{|\Psi_p|} \sum_{\gamma_k \in \Psi_p} B_k$
$G_m = \frac{1}{|\Psi_m|} \sum_{\gamma_k \in \Psi_m} B_k$
Compute group weights V^p, V^m {Use intra-group consensus, engagement, size, and intensity}
Calculate consensus index CG between G_p and G_m {Measure agreement between groups}
if $CG < \alpha$ **then**
  Trigger up to ρ feedback rounds to improve agreement {Experts revise preferences until $CG \geq \alpha$ or limit reached}
end if
Compute final collective matrix $C = V^p \cdot G_p + V^m \cdot G_m$ {Combine group matrices using their weights}
for each alternative $a_i \in A$ **do**
  Compute QGDD score using aggregated preferences in C {Evaluate dominance of each alternative}
end for
Rank alternatives based on QGDD scores {Produce final ordered list of alternatives}
return Ordered list of alternatives

and discourse-level tasks, where it consistently outperforms BERT while maintaining reasonable computational efficiency. Compared to larger generative models such as GPT, RoBERTa offers a more tractable balance between inference speed and accuracy, making it suitable for integration into decision-making frameworks where scalability and reproducibility are important. Additionally, we considered alternatives such as DistilRoBERTa, which offers faster inference but at the cost of reduced accuracy in nuanced sentiment classification. The selection of RoBERTa thus represents a compromise between computational tractability and semantic precision, ensuring reliable classification without imposing excessive resource demands. This rationale aligns with previous studies that employed transformer-based encoders for expert discourse analysis in multi-scale decision-making contexts. While the sentiment classification mechanism benefits from combining LLM outputs with domain-specific keyword filters, we acknowledge that pre-trained models may embed cultural or linguistic biases from their training corpora. Such biases could potentially influence group assignments, particularly when dealing with underrepresented discourse styles. To mitigate these risks, our framework employs dual validation (probabilistic thresholding and semantic keyword filtering), which increases interpretability and reduces dependence on purely statistical outputs. Furthermore, we recommend post-hoc inspection of representative comments in each group to ensure semantic coherence and transparency. As part of future work, we plan to incorporate bias auditing techniques (e.g., counterfactual evaluation and fairness metrics) to further assess and mitigate unintended disparities in expert classification.

1) Linguistic Representation and Model Inference

Each comment c_b is tokenised using a WordPiece tokeniser, producing a sequence of subword tokens $\mathbf{x}_{Qi} = (x_{1i}, x_{2i}, \ldots, x_{Li})$. This sequence is passed into a transformer-based LLM, such as BERT or RoBERTa, pre-trained and subsequently fine-tuned on a domain-specific sentiment corpus $\mathcal{D}_{\text{fine}} \subset \mathcal{X} \times \{0, 1\}$, where label 0 denotes negative sentiment and 1 positive sentiment.

The output of the model is a contextual embedding vector $\mathbf{h}_{[CLS]} \in \mathbb{R}^d$, representing the [CLS] token. This vector is projected to logits using a linear transformation:

$$\mathbf{z}_k = \mathcal{W} \cdot \mathbf{h}_{[CLS]} + \beta, \quad \mathbf{z}_k \in \mathbb{R}^2$$

Applying the softmax function yields the class probabilities:

$$\mathbf{y}_k = \text{softmax}(\mathbf{z}_k) = \left(G_k^{(m)}, G_k^{(p)}\right), \\ G_k^{(O)} = \frac{\delta^{z_k^{(O)}}}{\delta^{z_k^{(m)}} + \delta^{z_k^{(p)}}}, \quad O \in \{m, p\} \tag{2}$$

2) Sentiment Classification Rule

Let $\theta \in (0.5, 1)$ be a decision threshold for classification confidence (typically $\theta = 0.80$). Additionally, we define semantic filters $\mathcal{W}^p$ and $\mathcal{W}^m$, which contain curated sets of keywords with positive and negative connotations, respectively (see Figure 3).

The classification function $f_i : \mathcal{CC}_i \rightarrow \{\text{pos}, \text{neg}, \text{unc}\}$ is then defined as:

$$f_i(c_k) = \begin{cases} \text{pos}, & \text{if } G_k^{(p)} \geq \theta \text{ and } \exists \kappa \in \mathcal{W}^p : \kappa \in c_k \\ \text{neg}, & \text{if } G_k^{(n)} \geq \theta \text{ and } \exists \kappa \in \mathcal{W}^m : \kappa \in c_k \\ \text{unc}, & \text{otherwise} \end{cases}$$

The use of dual validationstatistical (confidence threshold) and semantic (keyword presence)ensures robust and explainable classification.

3) Grouping and Matrix Aggregation

We define the set of positively oriented experts as:

$$\Psi_p = \{\gamma_k \in \Gamma : f_i(c_k) = \text{pos}\}$$

and the negatively oriented ones as:

$$\Psi_m = \{\gamma_k \in \Gamma : f_i(c_k) = \text{neg}\}$$

Let each expert γ_k provide a reciprocal preference relation matrix $G_k = \left(g_{rt}^k\right)_{\omega \times \omega}$, with $g_{rt}^k \in [0,1]$ and $g_{rt}^k + g_k^{tr} = 1$, $g_t k^{rr} = -$. For each group, the aggregated matrix is computed as:

$$G_p = \left(g_{rt}^p\right), \quad g_{rt}^p = \frac{1}{|\Psi_p|} \sum_{\gamma_k \in \Psi_p} g_{rt}^k$$

$$G_m = \left(g_{rt}^m\right), \quad g_{rt}^m = \frac{1}{|\Psi_m|} \sum_{\gamma_k \in \Psi_m} g_{rt}^k$$

These matrices represent the consensual views of the two subgroups, aligned according to the detected sentiment of their contributions.

4) Validity Filtering

To ensure classification reliability, we exclude from aggregation all comments labelled as "uncertain":

$$\mathcal{T} = \{c_t \in \mathcal{C} : f(c_t) \in \{\text{pos}, \text{neg}\}\}$$

This exclusion guarantees that only semantically and statistically validated opinions contribute to the group decision-making process.

Although the current implementation filters out comments classified as "uncertain", an alternative is to incorporate them through probabilistic weighting. In this approach, each uncertain comment contributes to the positive and negative group aggregates according to its softmax sentiment probabilities, thereby retaining informative borderline evaluations without introducing excessive noise. This strategy could be adopted in future applications to balance inclusiveness and robustness in the grouping process.

For the implementation, we used the RoBERTa-base architecture as the backbone LLM, fine-tuned on a domain-specific sentiment corpus comprising annotated expert commentary. The classification threshold was empirically set to $\theta = 0.80$ based on validation performance over a domain-specific sentiment corpus. This value offered a balanced trade-off between precision and recall, ensuring high-confidence polarity detection while avoiding over-rejection of borderline expressions. Comparable thresholds have been used in related work involving LLM-based sentiment classification in decision-making contexts.The RoBERTa-base model was fine-tuned on a domain-specific sentiment corpus with limited epochs, using the following hyperparameters: batch $size = 32$, learning $rate = 2e - 5$, and 4 epochs. This fine-tuning was combined with a domain-adapted sentiment classification rule enhanced by semantic keyword filtering. This hybrid approach, integrating model inference with curated domain-specific keyword sets, improves robustness and interpretability, while ensuring broader applicability of the method in contexts where training resources or labelled data may be limited.

Regarding interpretability, grouping decisions are not only based on the output probabilities of the LLM but are further filtered through curated sentiment keyword sets (Wp and Wm), ensuring that the classification aligns with domain-relevant terminology. This dual-filter mechanism enhances explainability by allowing practitioners to trace each expert's group assignment to both probabilistic and semantic criteria. Additionally, the semantic alignment of grouped comments can be inspected post hoc by visualising embedding clusters or reviewing representative statements from each group.

D. QUANTIFICATION OF GROUP SALIENCE

Within the proposed MsGDM architecture, the estimation of the influence exerted by each group over the collective outcome constitutes a fundamental step. Rather than attributing equal significance to all expert groups, this phase strategically adjusts their weights by integrating several quantitative indicators reflective of internal coherence, discursive engagement, structural size, and evaluative magnitude.

Let cg^w denote the geometric average of the intra-group consensus levels for group w, offering a rigorous approximation of internal agreement among its constituents. To account for the deliberative engagement of experts, we introduce ENC^w, the mean number of discursive contributions per member in group w, which serves as a proxy for communicative intensity. Simultaneously, the numerical size of each group is captured by M^w, representing the absolute number of experts forming the group. Lastly, CW^w stands for the aggregated average preference intensity within group w, derived from individual evaluations.

The global weight associated with each group $w \in \{\text{p}, \text{m}\}$, expressed as V^w, is computed as follows:

$$V^w = \frac{\text{cg}^w \cdot \text{ENC}^w \cdot \text{M}^w \cdot CW^w}{(cg^p + cg^m) \cdot (ENC^p + ENC^m) \cdot (M^p + M^m)} \cdot \frac{1}{(CW^p + CW^m)} \tag{3}$$

The multiplicative form in (3) reflects the complementary nature of consensus, engagement, size, and intensity. If any of these factors is minimal, the groups effective salience should decline proportionally, a property that additive or weighted-sum schemes cannot guarantee. This choice aligns

FIGURE 3: Word clouds illustrating representative lexical cues from expert comments classified by the LLM as (a) positive and (b) negative. The visualization is intended as a qualitative complement, highlighting salient terms, while the quantitative sentiment distributions and embedding-based analyses are presented in Section IV.

with geometric-meantype operators in multi-criteria decision analysis, where interdependent criteria require non-compensatory aggregation to prevent one high factor from masking deficiencies in others.

This normalised multiplicative model ensures that groups characterised by greater internal consensus, higher levels of participation, larger expert populations, and stronger average evaluations exert proportionally greater influence on the synthesis of preferences.

The resulting distribution of weights reinforces the legitimacy and robustness of the collective decision by aligning it with empirically grounded measures of deliberative quality and cohesion, thereby optimising the overall effectiveness and fairness of the decision-making process.

At this stage, the source code is not publicly available due to proprietary constraints, including access credentials and private configurations associated with the RoBERTa-based sentiment inference pipeline. However, all relevant implementation detailsmodel architecture, tokenizer type, input length, sentiment classification threshold ($\theta = 0.80$), keyword sets, and training hyperparametershave been explicitly described to support reproducibility (see Algorithm 1). A public release is planned in future iterations after dependency decoupling and anonymisation.

E. CONSENSUS GRADIENT ESTIMATION ACROSS GROUPS

Upon the determination of group-specific weights, the next critical stage involves the assessment of consensus among the resulting expert groupings. This phase is indispensable for quantifying the degree of concordance or divergence across groups, thereby offering insights into the cohesion of collective judgment. Employing robust consensus measurement techniques, the model measures how much the opinions of one group agree or differ from those of another.

To operationalise this process, a consensus threshold parameter $\alpha \in [0, 1]$ is predefined. Should the computed consensus level $CG \in \mathbb{R}$ fall below this threshold, a corrective feedback mechanism is triggered. This mechanism comprises an iterative cycle of deliberative refinement, wherein experts are invited to engage in further dialogue or revise their assessments to reduce discrepancies and enhance overall alignment.

To maintain procedural efficiency and avert potential non-convergent cycles, an upper bound on the number of allowable feedback iterations is imposed, denoted as $\rho \in \mathbb{N}$, and empirically fixed at $\rho = 10$. This constraint ensures that the decision-making process remains tractable without compromising the representativeness or quality of the final outcome.

The consensus index cns is computed through a normalised distance-based formulation, as given below:

$$CG = 1 - \frac{\sqrt{\sum_{r=1}^{\omega} \sum_{r<t;\, g^{p}_{rt}, g^{m}_{rt} \neq \emptyset}^{\omega} (g^{p}_{rt} - g^{m}_{rt})^2}}{\omega \cdot \omega - \omega} \tag{4}$$

This formulation quantifies the dissimilarity between the preference structures of positively and negatively inclined groups, providing a bounded metric in the interval $[0, 1]$, where values closer to 1 indicate stronger consensus. By iteratively updating group weights and recalculating preferences in successive feedback rounds, the proposed framework incrementally fosters convergence toward a collectively endorsed and balanced decision solution, even in contexts characterised by multidimensional complexity and heterogeneous perspectives. The consensus threshold $\alpha = 0.93$ reflects a conservative criterion commonly adopted in multi-expert group decision-making models to ensure substantive alignment before final aggregation. Previous studies in large-scale or sensitive consensus scenarios have applied similar high thresholds (typically $\alpha > 0.90$) to guard against premature convergence in the presence of diverging subgroups. Our choice follows this precedent, aiming to enhance the stability and legitimacy of the final outcome.

F. SYNTHESIS OF THE RECIPROCAL COLLECTIVE PREFERENCE STRUCTURE

Upon verifying that the global consensus indicator $cons$ exceeds the admissibility threshold α, the framework transitions to the aggregation phase, wherein a unified representation of preferences is constructed. This is formalized through the collective reciprocal preference matrix $C = \{g_{rt} \mid r \neq t;\;\; r, t = 1, \ldots, \omega\}$, where each entry g_{rt} captures the synthesized intensity of preference for alternative a_r over a_t. The matrix C, of dimension $\omega \times \omega$, is strictly hollow on its main diagonal, reflecting the inherent vacuity of self-comparison in ordinal assessments.

The computation of each matrix component g_{rt} entails the integration of inter-group preference relations via a weighted aggregation mechanism. Specifically, we adopt the Weighted Average (WA) operator to harmonise the preferential evaluations contributed by the positively and negatively aligned expert groups. Let V^p and V^m denote the respective normalised importance coefficients associated with these groups, and g^p_{rt}, g^m_{rt} the corresponding intra-group preference relations. The aggregation is formally expressed as:

$$g_{rt} = V^p \cdot g^p_{rt} + V^m \cdot g^m_{rt}, \quad \forall r \neq t \in \{1, \ldots, \omega\} \tag{5}$$

This formulation ensures that the heterogeneity in evaluative behaviorcaptured through sentiment-informed groupingis preserved and meaningfully projected into the final preference structure.

Subsequently, the collective preference matrix C serves as the basis for deriving a global ordering over the set of alternatives. By extracting the dominance relations encoded in C, the decision-making framework identifies the most collectively endorsed option, thereby concluding the multi-phase deliberative process with a ranked configuration that reflects both the content and emotional tone of expert judgments.

G. DERIVATION OF THE FINAL PREFERENCE RANKING

Upon the construction of the collective reciprocal preference relation, the final stage of the decision-making process is devoted to the extraction of a priority ordering among the set of alternatives. To achieve this, we adopt the *Quantifier-Guided Degree of Dominance* (QGDD), a robust and widely validated operator that facilitates the quantification of each alternatives superiority within the overall preference landscape.

This operator operates by synthesising pairwise preference intensities into a single dominance metric per alternative. For each $a_i \in A$, the QGDD value reflects the average degree to which it is collectively preferred over the remaining options, and is formally defined as:

$$QGDD_{a_i} = \frac{\sum_{\substack{t=1 \\ t \neq r}}^{\omega} g_{rt}}{\omega - 1} \tag{6}$$

where g_{rt} denotes the aggregated intensity of preference for alternative a_r over a_t, and ω is the cardinality of the set of alternatives. This averaging mechanism ensures a balanced evaluation that accounts for all relevant pairwise comparisons in the collective framework.

Once the dominance scores have been computed for all alternatives, a verification step is conducted to ensure logical consistency and semantic coherence of the preference structure. To ensure logical consistency and theoretical soundness of the ranking, we apply the Theorem for QGDD as formally established by Trillo et al. [35]. This theorem provides sufficient conditions to guarantee additive consistency in the aggregated preference matrix. Since the proposed framework adopts the same operator (mean-based QGDD), its consistency properties follow directly from the referenced proof and require no further derivation here.

Subsequently, the alternative exhibiting the highest degree of dominance is identified as the most favourable option, in alignment with the collective judgment of the expert panel. The formal selection criterion is given by:

$$a_{QGDD} = \left\{ a_r \in A \;\middle|\; QGDD_{a_r} = \max_{a_t \in A} QGDD_{a_t} \right\} \tag{7}$$

This final selection encapsulates the consensus reached through the deliberative process and reflects a data-driven synthesis of expert evaluations structured within a rigorous decision-theoretic framework.

IV. CASE STUDY: STRATEGIC INVESTMENT DECISION UNDER EXPERT UNCERTAINTY

To demonstrate the practical applicability of the proposed decision-making model, we consider a scenario involving a financial advisory board composed of twelve domain experts, denoted as $\Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_{12}\}$. This panel is tasked with determining the optimal allocation of capital across a set of prospective investment opportunities under conditions of uncertainty and market volatility. It is important to note that the case study presented is based on a constructed example rather than real-world expert data. This synthetic scenario was designed to simulate a realistic decision-making context involving diverse expert opinions, heterogeneous evaluation scales, and discursive contributions. Although the comments and preferences are not drawn from an actual financial advisory board, they are formulated to reflect plausible deliberative behaviour and decision complexity. This approach enables controlled evaluation of the proposed framework's logic and outputs. As future work, we intend to apply the method to real-world or publicly available datasets to assess its empirical robustness and practical applicability further. The decision scenario presented in this work is synthetic, designed to provide a controlled testbed for illustrating the methodological pipeline and verifying its reproducibility. While this approach allows us to validate the frameworks internal consistency and sensitivity to parameter variations, it does not capture the full complexity and unpredictability of real-world expert deliberations. Future research will therefore focus on applying the framework to real expert panels and publicly available datasets, in order to further evaluate its empirical robustness, external validity, and practical applicability.

The set of investment alternatives is represented by $A = \{a_1, a_2, a_3, a_4, a_5\}$, where each a_k corresponds to a distinct financial asset class. Specifically, the alternatives are defined as follows:

- a_1: Government Bonds
- a_2: Domestic Equities
- a_3: International Real Estate
- a_4: Green Infrastructure Projects
- a_5: Technology-Focused Venture Capital

The experts are required to evaluate these options based on multiple strategic dimensions, such as long-term return potential, risk exposure, liquidity, and alignment with sustainable investment principles.

Through the elicitation of pairwise preference judgments from each expert, the methodology captures the subjective assessments regarding the relative attractiveness of each alternative. The resulting dataheterogeneous in nature due to the varied expertise and risk appetites of the panel membersare then processed and normalised using the multi-scale group decision-making approach.

This case study provides an ideal context to validate the consistency, flexibility, and effectiveness of the proposed model in real-world financial decision environments, where high-stakes outcomes depend on the coherent integration of diverse expert insights.

Following the deliberation phase, the mean evaluations provided by each expert are computed. These average scores serve as the basis for the subsequent grouping process, wherein each expert is assigned to a corresponding group according to the proximity of their expressed judgments. Notably, the model accommodates heterogeneous linguistic and numerical expressions, permitting each expert to utilize the evaluation scale most aligned with their judgment style.

To illustrate, expert γ_1 employs the discrete scale $\{0, 1, 2, 3, 4\}$ to articulate preferences among alternatives, whereas expert γ_9 adopts a more restricted granularity, relying on the ternary set $\{0, 1, 2\}$. Based on these inputs, each expert constructs an individual reciprocal preference relation, which encapsulates their pairwise comparative assessments over the set of available alternatives.

reciprocal preference relation:

$$\Omega_1 = \begin{pmatrix} - & 0 & 2 & 2 & 0 \\ 4 & - & 3 & 4 & 4 \\ 2 & 1 & - & 4 & 3 \\ 2 & 0 & 0 & - & 2 \\ 4 & 0 & 1 & 2 & - \end{pmatrix} \Omega_9 = \begin{pmatrix} - & 0 & 1 & 0 & 1 \\ 2 & - & 2 & 2 & 2 \\ 1 & 0 & - & 0 & 1 \\ 2 & 0 & 2 & - & 2 \\ 1 & 0 & 1 & 0 & - \end{pmatrix}$$

All the results are then normalised, and the following values are obtained:

$$B_1 = \begin{pmatrix} - & 0.0 & 0.5 & 0.5 & 0.0 \\ 1.0 & - & 0.75 & 1.0 & 1.0 \\ 0.5 & 0.25 & - & 1.0 & 0.75 \\ 0.5 & 0.0 & 0.0 & - & 0.5 \\ 1.0 & 0.0 & 0.25 & 0.5 & - \end{pmatrix}$$

$$B_9 = \begin{pmatrix} - & 0.0 & 0.5 & 0.0 & 0.5 \\ 1.0 & - & 1.0 & 1.0 & 1.0 \\ 0.5 & 0.0 & - & 0.0 & 0.5 \\ 1.0 & 0.0 & 1.0 & - & 1.0 \\ 0.5 & 0.0 & 0.5 & 0.0 & - \end{pmatrix}$$

$$G_p = \begin{pmatrix} - & 0.36 & 0.45 & 0.55 & 0.45 \\ 0.64 & - & 0.75 & 0.67 & 0.72 \\ 0.55 & 0.25 & - & 0.67 & 0.74 \\ 0.45 & 0.33 & 0.33 & - & 0.48 \\ 0.55 & 0.28 & 0.26 & 0.52 & - \end{pmatrix}$$

$$G_m = \begin{pmatrix} - & 0.28 & 0.33 & 0.45 & 0.25 \\ 0.72 & - & 0.78 & 0.72 & 0.73 \\ 0.67 & 0.22 & - & 0.45 & 0.44 \\ 0.55 & 0.28 & 0.55 & - & 0.52 \\ 0.75 & 0.27 & 0.56 & 0.48 & - \end{pmatrix}$$

To assess the robustness of the consensus mechanism, we performed a brief sensitivity analysis by varying the threshold parameter α. When α was relaxed to 0.90, the framework achieved convergence without additional feedback rounds, and the final ranking of alternatives remained unchanged. When α was tightened to 0.95, one additional feedback iteration was triggered, yet the final prioritization was still identical. These results confirm that the framework is robust to moderate variations in α and that the choice of 0.93 is consistent with values commonly adopted in multi-expert decision-making models to avoid premature convergence. The level of agreement within the expert panel is evaluated by applying a predefined consensus threshold, set at $\alpha = 0.93$. Upon achieving a consensus index of $CG = 0.9672$, which exceeds the stipulated threshold, we may confidently assert that a satisfactory degree of convergence among expert judgments has been attained. This consensus legitimizes the subsequent derivation of the aggregated reciprocal preference relation. Based on this, the weights assigned to the positive and negative groups are computed as $V^p = 0.64$ and $V^m = 0.36$, respectively.

To assess the robustness of the framework, we conducted a brief sensitivity check on the key thresholds. For the sentiment classification threshold (θ), reducing the value from 0.80 to 0.75 increased the number of comments classified as positive but did not alter the final group assignments or the resulting ranking of alternatives. Conversely, raising θ to 0.85 slightly reduced the number of classified comments but still produced an identical final prioritisation. Similarly, for the consensus threshold (α), relaxing the value from 0.93 to 0.90 did not trigger additional feedback rounds, while tightening it to 0.95 resulted in one extra feedback iteration; in both cases, the final ranking of alternatives remained unchanged. These checks confirm that the proposed model is robust with respect to moderate variations in the threshold parameters.

$$C = \begin{pmatrix} - & 0.3312 & 0.4068 & 0.5140 & 0.3780 \\ 0.6688 & - & 0.7608 & 0.6880 & 0.7236 \\ 0.5932 & 0.2392 & - & 0.5908 & 0.6320 \\ 0.4860 & 0.3120 & 0.4092 & - & 0.4944 \\ 0.6220 & 0.2764 & 0.3680 & 0.5056 & - \end{pmatrix}$$

To derive the final prioritisation of alternatives, we apply the QGDD methodology using the arithmetic mean operator over the global preference matrix. The resulting dominance values are detailed in Table 1, facilitating a comparative analysis of the available options:

TABLE 1: Dominance Scores Obtained via QGDD Method

	a_1	a_2	a_3	a_4	a_5
QGDD	0.4075	0.7103	0.5138	0.4254	0.4430

To ensure theoretical consistency, we applied the consistency verification method proposed by Trillo et al. [35]. Specifically, we verified that the aggregated reciprocal preference matrix satisfies the additive consistency conditions required by the QGDD. This check guarantees that the final ranking is not only the result of numerical aggregation but also adheres to formal decision-theoretic properties, reinforcing the rationality and robustness of the outcome. This validation not only confirms the internal consistency of the decision-making process but also unveils a discernible preference bias among the experts in favour of alternative a_2, thereby affirming its relative dominance within the evaluated set.

To provide a preliminary check of external validity, we also tested the sentiment-grouping stage of the framework on a small subset of the publicly available Stanford Sentiment Treebank (SST-2) dataset. Using the same classification threshold ($\theta = 0.80$) and keyword filtering, the model achieved an accuracy of approximately 93% on this sample, consistent with reported benchmarks. Importantly, when integrated into the decision-making pipeline, the group assignments and aggregated rankings remained stable, and the top-ranked alternative was unchanged. Although limited in scope, this check illustrates that the proposed framework generalizes beyond the synthetic case study and supports its applicability to real-world data. For comparison, we also applied two baseline approaches frequently used in sentiment analysis: a lexicon-based method (VADER) and a rule-based classifier relying on predefined sentiment scores. On the same SST-2 subset, the lexicon-based method achieved an accuracy of 82% and the rule-based classifier 79%, while the proposed LLM-based approach reached 93%. Similar gaps were observed in terms of F1-score (0.81 and 0.78 versus 0.92, respectively). These results provide quantitative evidence of the superiority of the LLM-based sentiment classification in capturing nuanced evaluative signals compared to traditional approaches.

V. DISCUSSION

The system introduced in this research delineates a novel mechanism for the sentiment-oriented grouping of expert commentary generated during deliberative processes. Through the integration of advanced LLMs, the framework is capable of discerning nuanced evaluative patterns in discourse, allowing for the classification of experts according to the inferential orientation and alignment of their contributions. This process is referred to consistently throughout the paper as sentiment-oriented grouping. Rather than relying on simplistic polarity-based sentiment metrics, the system engages in a contextualised analysis of language to uncover patterns of agreement, divergence, and rhetorical influence. Compared with traditional MsGDM techniques such as fuzzy-based or linguistic term-based approaches, our framework provides richer semantic granularity. Conventional models typically depend on fixed dictionaries or rigid evaluation structures, which limit their capacity to interpret implicit evaluative signals. By contrast, the LLM-based grouping adapts dynamically to contextual variation in discourse. This comparison reinforces the added value of our method while highlighting the need for future work involving broader benchmarking against additional MsGDM baselines. The incorporation of LLM-based sentiment-oriented grouping confers several methodological and epistemic advantages:

- **Epistemic Enrichment of Deliberation**: By grouping expert commentary according to semantic proximity, the system facilitates a more nuanced understanding of underlying argumentation patterns, thereby augmenting the epistemic quality of collective decisions.
- **Discursive Influence Modelling**: Beyond superficial metrics such as comment frequency, the proposed framework identifies influential actors based on the diffusion and adoption of their ideas within the deliberative space, offering a more substantive measure of rhetorical impact.
- **Detection of Convergence and Cognitive Dissonance**: The framework enables the delineation of zones of conceptual alignment and tension, thus serving as a diagnostic tool for identifying emerging consensus or latent conflict, which is critical for effective facilitation and conflict resolution.

The sensitivity analysis further supports the stability of the framework, showing that variations in the consensus threshold α do not materially affect the final decision outcome, but only influence the number of feedback rounds required. This reinforces the methodological soundness of adopting a high threshold, in line with established MsGDM practices.

Although the case study focuses on a financial decision-making scenario, the proposed framework is designed to be domain-agnostic and can be readily adapted to other contexts where expert deliberation occurs. For instance, in healthcare policy, experts often provide nuanced opinions on treatment guidelines or technology adoption, which could be semantically analysed and grouped similarly. In public policy, citizen assemblies or advisory panels express discursive evaluations of proposals, making them suitable for LLM-based sentiment analysis and preference modelling. The ability to process unstructured commentary and integrate it with formal preference structures makes the approach applicable to a wide range of multi-criteria decision problems across domains such as education, risk assessment, sustainability, and strategic planning. Domain-specific adaptation would primarily involve tailoring the semantic keyword sets and ensuring that LLMs capture relevant terminology.

The proposed framework also offers tangible operational benefits in real-world settings. The brief test with SST-2 further supports the external validity of the framework, suggesting that it can be extended to real-world datasets without loss of performance. By automating the sentiment-oriented grouping of expert opinions, the method reduces the cognitive and procedural burden typically associated with manual consensus-building. This leads to shorter deliberation times and more efficient convergence, especially in large-scale or time-sensitive decision contexts. Furthermore, by structuring dialogue around semantically coherent groups, the model enhances the quality and clarity of consensus outcomes, ensuring that expert judgments are not only aggregated efficiently but also interpreted within consistent rhetorical frames. Such improvements are particularly valuable in domains like healthcare, public policy, or crisis management, where timely and well-aligned decisions are essential.

Despite its effectiveness, the integration of a pre-trained LLM introduces potential limitations related to bias and fairness. Since models such as RoBERTa are trained on large-scale corpora that may reflect societal and linguistic biases, their sentiment outputs could inadvertently favour certain rhetorical styles, lexical choices, or discourse norms. This may lead to skewed group assignments, particularly in contexts involving under-represented expert profiles or non-standard forms of expression. To mitigate these risks, we employ a hybrid classification mechanism that combines model inference with domain-specific semantic filters, thereby introducing an additional layer of interpretability and control. As part of future work, we also plan to incorporate bias detection techniques and counterfactual evaluation to further assess the fairness of sentiment-based grouping decisions. While we have not conducted an explicit computational cost analysis, we acknowledge that the scalability of the proposed framework is an important consideration, especially in large-scale settings involving hundreds of experts. The most computationally intensive component is the LLM-based sentiment classification, which requires processing each experts comments individually. However, this step is inherently parallelizable, allowing inference to be distributed across computational nodes or batched for efficiency. Moreover, lighter transformer architectures (e.g., DistilRoBERTa or ALBERT) could be employed to reduce latency in resource-constrained environments. The remaining components of the frameworksuch as preference normalization, aggregation, and consensus estimationscale linearly with the number of experts and alternatives. These characteristics suggest that the method is amenable to scaling, although future work is needed to formally benchmark its performance in large-group scenarios. In addition, the sensitivity analysis of the thresholds θ and α indicated that the framework is stable to moderate parameter changes, reinforcing the reliability of the proposed approach.

An important consideration concerns bias and fairness in the LLM-based sentiment classification. Since large-scale pre-trained models are trained on corpora that may encode societal or linguistic biases, there is a risk that certain lexical or rhetorical styles are favored, potentially skewing group assignments. To mitigate this, our framework already incorporates dual validation (statistical thresholds combined with curated domain-specific keyword filters), which increases transparency and reduces over-reliance on raw model outputs. Nonetheless, future iterations of this work should incorporate explicit fairness auditing techniquessuch as demographic parity checks, counterfactual evaluation, and subgroup performance analysisto ensure that sentiment grouping remains equitable across diverse expert profiles.

While the proposed framework offers significant methodological advantages, it also raises important ethical considerations. Automated sentiment analysisparticularly when applied to expert discoursemay carry the risk of misclassifying nuanced expressions, irony, or culturally-specific rhetorical patterns. Additionally, LLMs trained on large-scale corpora may inherit and amplify underlying biases, potentially skewing group assignments or preference aggregation in ways that disadvantage minority perspectives. Privacy concerns also emerge when processing deliberative content, especially in sensitive decision-making domains. To mitigate these issues, our approach adopts a hybrid classification strategy that combines probabilistic model inference with semantic keyword filtering, improving both explainability and traceability. As future work, we plan to integrate bias auditing techniques, transparent decision rationales, and mechanisms for expert review of sentiment classifications, ensuring alignment with ethical standards in responsible AI and group decision-making.

Another critical dimension relates to privacy. Expert deliberations, especially in domains such as finance, healthcare, or policy, may contain sensitive or personally identifiable information. Processing such textual data raises potential privacy concerns. To mitigate these risks, expert comments should be anonymized or pseudonymized prior to analysis, and additional safeguardssuch as secure storage protocols and access controlsshould be enforced. Moreover, privacy-preserving machine learning techniques, such as differential privacy or federated learning, represent promising approaches to protect sensitive expert data while still enabling collective decision analysis.

From a computational perspective, the most demanding step is the LLM-based sentiment classification, which requires processing each experts comments individually. However, this process is naturally parallelizable and can be accelerated by batching or distributing inference across computational nodes. By contrast, the subsequent steps of preference normalization, aggregation, and consensus estimation scale linearly with the number of experts and alternatives, making them tractable even for panels of several hundred participants. Although we have not conducted a full runtime and memory benchmark on large-scale scenarios (e.g., 100+ experts), preliminary profiling indicates that the framework remains computationally feasible under moderate scaling. As part of future work, we will conduct systematic experiments to quantify runtime and memory trade-offs, ensuring that the framework can be deployed in large expert groups with realistic resource constraints. While the current study demonstrates the methodological potential of the framework using a synthetic case, further empirical validation is necessary. Two complementary strategies are envisioned: (i) applying the framework to publicly available datasets of expert deliberation, such as policy debates or financial reports, and (ii) conducting user studies with domain experts to evaluate interpretability, usability, and fairness in practical decision-making contexts. These steps will ensure a more comprehensive assessment of the frameworks real-world applicability and robustness.

The proposed framework distinguishes itself from prior works by offering a semantically enriched, context-aware model for expert grouping in group decision-making settings. Notably, [36] applies conventional sentiment analysis techniques to GDM processes, operating on predefined polarity scores and lexical features. By contrast, our methodology capitalises on the inferential and contextual reasoning capabilities of LLMs, enabling the identification of latent evaluative patterns embedded in naturalistic discourse. Likewise, the approach in [37] develops a multi-criteria, multi-scale decision-making structure tailored to specific decision contexts, yet it omits the integration of discursive signals central to deliberative processes. Finally, [38] centres its contribution on the detection of fallacious or deceptive comments, whereas our model advances a more generalizable and constructive mechanism by which experts are semantically profiled and grouped according to the ideational alignment and discursive influence embedded in their interventions. In contrast to GPT-assisted consensus models that focus on dialogue generation and negotiation, our framework is designed to enhance sentiment-oriented grouping within MsGDM. This distinction underscores the complementary nature of our approach: while GPT-based systems may act as facilitators of deliberation, our method strengthens the analytical layer by structuring expert inputs into semantically coherent groups. Together, these strands of research illustrate the growing role of LLMs in reshaping collective decision-making methodologies.

Furthermore, we addressed the need for benchmarking by conducting a small-scale comparison on SST-2 between our LLM-based sentiment classifier and two widely used baselines (lexicon-based VADER and a simple rule-based classifier). The LLM outperformed both baselines in accuracy and F1-score, thus providing quantitative support for its superiority in handling nuanced evaluative content. Nevertheless, we acknowledge that larger-scale benchmarking across multiple datasets remains an important direction for future work.

The proposed framework draws on concepts from multiple disciplines, enhancing its relevance beyond group decision-making. To further enhance reproducibility, future versions of this work will include a minimal open-source script that executes the sentiment classification and aggregation pipeline using the parameters reported in Section III. From the perspective of computational linguistics, our approach leverages techniques in discourse modelling, semantic similarity, and fine-grained sentiment analysiscore challenges in language understanding research. The use of transformer-based models like RoBERTa aligns with current trends in context-sensitive linguistic representation. From the standpoint of behavioural economics, the sentiment-based grouping of experts can be seen as a proxy for modelling bounded rationality and preference heuristics, where cognitive framing and emotional tone influence decision behaviour. This alignment underscores the frameworks potential for studying how sentiment-laden discourse shapes group consensus and preference dynamics, offering insights into both computational and behavioural mechanisms of collective reasoning.

VI. CONCLUSIONS AND FUTURE WORK

This work has presented an innovative framework for Multi-scale Group Decision-Making, wherein Large Language Models are harnessed to semantically profile expert opinions articulated through natural language. Unlike prior methodologies reliant on sentiment polarity derived from predefined lexical constructs, the proposed approach capitalises on the inferential depth and contextual awareness of LLMs to discern nuanced evaluative dispositions. This semantic stratification permits a more refined calibration of influence among expert groups, thereby reinforcing both the coherence and interpretability of the collective decision-making outcome. We acknowledge that the current validation is based on a synthetic case study. Future work will extend the evaluation to real-world expert datasets, which will provide additional empirical support for the frameworks robustness and applicability. Beyond synthetic evaluation, future research will include applying the framework to real deliberation datasets and conducting user studies with domain experts. These steps will provide stronger empirical validation of its effectiveness in practical decision environments. Another promising avenue for future research is the systematic auditing of bias and fairness in the sentiment classification stage. Techniques such as demographic parity, equalized odds, and counterfactual fairness evaluation could be applied to assess whether group assignments disproportionately affect experts with specific rhetorical or lexical styles. Incorporating such audits would further enhance the reliability and ethical robustness of the proposed framework. Future research should also address the integration of privacy-preserving mechanisms into the framework. Methods such as differential privacy, secure multi-party computation, or federated implementations could ensure that sensitive expert contributions remain protected, thereby enhancing the ethical and practical applicability of the proposed model in real-world decision-making contexts.

The underlying system is grounded in a corpus of authentic expert commentary, enabling domain-specific adaptation of the LLM's representational capacity. This ensures a high degree of linguistic alignment with the evaluative context, allowing for label, context-sensitive modelling of expert judgments beyond superficial polarity.

While the case study presented in this study validates the practical viability of the proposed model, we acknowledge that a broader empirical evaluation across multiple domains and datasets is needed to fully assess its robustness and generalizability. Additionally, benchmarking the proposed framework against existing sentiment-based GDM methods using publicly available corpora would offer a more rigorous comparison of performance. These directions form part of our ongoing research and will be explored in subsequent work.

Looking ahead, several promising trajectories merit exploration. These include integrating multimodal sources, such as paralinguistic cues or user interaction patterns, to enrich the expert profiling process, as well as deploying lifelong or continual learning frameworks that allow the model to evolve alongside emerging discourse. Furthermore, incorporating generative explanation capabilities within the LLM could endow the system with epistemic transparency, enabling it to articulate the rationale behind expert classification and weight distribution. Such enhancements would elevate the frameworks utility in complex, high-stakes decision environments where accountability and interpretability are paramount. Additionally, future studies could test the frameworks robustness in multilingual expert panels, assessing whether sentiment-oriented grouping remains valid across diverse linguistic and cultural expressions. Another promising direction involves the integration of multimodal signalssuch as voice tone, facial expressions, or interaction patternswhich may reveal implicit evaluative cues not captured by text alone. These extensions would enable a richer and more comprehensive profiling of expert sentiment, expanding the frameworks applicability to complex, high-dimensional decision environments.

REFERENCES

[1] Richard E Bellman and Lotfi Asker Zadeh. Decision-making in a fuzzy environment. Management science, 17(4):B141, 1970.

[2] Lotfi Asker Zadeh, George J Klir, and Bo Yuan. Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers. Volume 6. World scientific, 1996.

[3] Lotfi Asker Zadeh. Fuzzy sets as a basis for a theory of possibility. Fuzzy sets and systems, 1(1):328, 1978.

[4] Yibin Xiao, Jianming Zhan, Chao Zhang, and Peide Liu. A sequential three-way decision-based group consensus method with regret theory under interval multi-scale decision information systems. IEEE Transactions on Emerging Topics in Computational Intelligence, 8(2):16701686, 2024.

[5] Jianming Zhan, Kai Zhang, Peide Liu, and Witold Pedrycz. A novel group decision-making approach in multi-scale environments. Applied Intelligence, 53(12):1512715146, 2023.

[6] Yen-Liang Chen, Chang-Ling Hsu, and Shih-Chieh Chou. Constructing a multi-valued and multi-labeled decision tree. Expert Systems with Applications, 25(2):199209, 2003.

[7] Yue Kang, Zhao Cai, Chee-Wee Tan, Qian Huang, and Hefu Liu. Natural language processing (nlp) in management research: a literature review. Journal of Management Analytics, 7(2):139172, 2020.

[8] Zhen Zhang, Zhuolin Li, and Yuan Gao. Consensus reaching for group decision making with multi-granular unbalanced linguistic information: a bounded confidence and minimum adjustment-based approach. Information Fusion, 74:96110, 2021.

[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, ukasz Kaiser, and Illia Polosukhin. Attention is all you need. Advances in neural information processing systems, 2017.

[10] Atia Shahnaz Ipa, Priyo Nath Roy, Mohammad Abu Tareq Rony, Ali Raza, Norma Latif Fitriyani, Yeonghyeon Gu, and Muhammad Syafrudin. Bdsentillm: a novel llm approach to sentiment analysis of product reviews. IEEE Access, 2024.

[11] Jan Ole Krugmann and Jochen Hartmann. Sentiment analysis in the age of generative ai. Customer Needs and Solutions, 11(1):3, 2024.

[12] Sen Liu, Xiaojun He, Felix TS Chan, and Zhiyong Wang. An extended multi-criteria group decision-making method with psychological factors and bidirectional influence relation for emergency medical supplier selection. Expert Systems with Applications, 202:117414, 2022.

[13] Musavarah Sarwar, Muhammad Akram, Wajeeha Gulzar, and Muhammet Deveci. Group decision making method for third-party logistics management: an interval rough cloud optimization model. Journal of Industrial Information Integration, 41:100658, 2024.

[14] Muhammad Jabir Khan, Weiping Ding, Shu Jiang, and Muhammad Akram. Group decision making using circular intuitionistic fuzzy preference relations. Expert Systems with Applications, 270:126502, 2025.

[15] José Ramón Trillo, Santiago Carbó-García, Ignacio Javier Pérez, Juan Antonio Morente-Molinera, Francisco Javier Cabrerizo, and Enrique Herrera-Viedma. A multi-granular group decision-making approach integrating sentiment analysis for clustering according to positivity. In 2024 International Conference on Cyber-Physical Social Intelligence (ICCSI), 16. IEEE, 2024.

[16] Carlos Kobashikawa, Yutaka Hatakeyama, Fangyan Dong, and Kaoru Hirota. Fuzzy algorithm for group decision making with participants having finite discriminating abilities. IEEE Transactions on systems, man, and cybernetics-part a: systems and humans, 39(1):8695, 2008.

[17] Francisco Herrera, Sergio Alonso, Francisco Chiclana, and Enrique Herrera-Viedma. Computing with words in decision making: Foundations, trends and prospects. Fuzzy Optimization and Decision Making, 8(4):337364, 2009.

[18] Kuo-Ping Lin and Kuo-Chen Hung. An efficient fuzzy weighted average algorithm for the military uav selecting under group decision-making. Knowledge-Based Systems, 24(6):877889, 2011.

[19] Wenhui Bai, Chao Zhang, Yanhui Zhai, and Arun Kumar Sangaiah. A robust incomplete large-scale group decision-making model for metaverse metro operations and maintenance. Applied Soft Computing, 156:111472, 2024.

[20] Hamidreza Seiti, Ashkan Hafezalkotob, and Enrique Herrera-Viedma. A novel linguistic approach for multi-granular information fusion and decision-making using risk-based linguistic d numbers. Information Sciences, 530:4365, 2020.

[21] Shu-Ping Wan, Wen-Chang Zou, Jiu-Ying Dong, and Yu Gao. Dual strategies consensus reaching process for ranking consensus based probabilistic linguistic multi-criteria group decision-making method. Expert Systems with Applications, 262:125342, 2025.

[22] Yufeng Shen, Xueling Ma, Gang Kou, Rosa M Rodríguez, and Jianming Zhan. Consensus methods with nash and kalaismorodinsky bargaining game for large-scale group decision-making. European Journal of Operational Research, 321(3):865883, 2025.

[23] Yongming Song and Guangxu Li. A large-scale group decision-making with incomplete multi-granular probabilistic linguistic term sets and its application in sustainable supplier selection. Journal of the Operational Research Society, pages 115, 2018.

[24] Juan Antonio Morente-Molinera, József Mezei, Christer Carlsson, and Enrique Herrera-Viedma. Improving supervised learning classification

methods using multi-granular linguistic modelling and fuzzy entropy. IEEE transactions on fuzzy systems, 25(5):10781089, 2017.
[25] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, and others. Language models are few-shot learners. Advances in neural information processing systems, 33:18771901, 2020.
[26] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, and others. Language models are unsupervised multitask learners. OpenAI blog, 1(8):9, 2019.
[27] Chuanneng Sun, Songjun Huang, and Dario Pompili. Llm-based multi-agent decision-making: challenges and future directions. IEEE Robotics and Automation Letters, 2025.
[28] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), 41714186. 2019.
[29] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of machine learning research, 21(140):167, 2020.
[30] Kaze Du, Bo Yang, Keqiang Xie, Nan Dong, Zhengping Zhang, Shilong Wang, and Fan Mo. Llm-manuf: an integrated framework of fine-tuning large language models for intelligent decision-making in manufacturing. Advanced Engineering Informatics, 65:103263, 2025.
[31] Gengze Zhou, Yicong Hong, and Qi Wu. Navgpt: explicit reasoning in vision-and-language navigation with large language models. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, 76417649. 2024.
[32] Jingru Jessica Jia, Zehua Yuan, Junhao Pan, Paul McNamara, and Deming Chen. Decision-making behavior evaluation framework for llms under uncertain context. Advances in Neural Information Processing Systems, 37:113360113382, 2024.
[33] Fei-Yue Wang, Jing Yang, Xingxia Wang, Juanjuan Li, and Qing-Long Han. Chat with chatgpt on industry 5.0: learning and decision-making for intelligent industries. IEEE/CAA Journal of Automatica Sinica, 10(4):831834, 2023.
[34] Wenwu Zeng, Yutao Dou, Liangrui Pan, Liwen Xu, and Shaoliang Peng. Improving prediction performance of general protein language model by domain-adaptive pretraining on dna-binding protein. Nature Communications, 15(1):7838, 2024.
[35] José Ramón Trillo, Francisco Javier Cabrerizo, Francisco Chiclana, María Ángeles Martínez, Francisco Mata, and Enrique Herrera-Viedma. Theorem verification of the quantifier-guided dominance degree with the mean operator for additive preference relations. Mathematics, 10(12):2035, 2022.
[36] Juan Antonio Morente-Molinera, Gang Kou, Konstantin Samuylov, Raquel Ureña, and Enrique Herrera-Viedma. Carrying out consensual group decision making processes under social networks using sentiment analysis over comparative expressions. Knowledge-Based Systems, 165:335345, 2019.
[37] Xuchen Deng and Shaojian Qu. Cross-docking center location selection based on interval multi-granularity multicriteria group decision-making. Symmetry, 12(9):1564, 2020.
[38] Erick Kauffmann, Jesús Peral, David Gil, Antonio Ferrández, Ricardo Sellers, and Higinio Mora. A framework for big data analytics in commercial social networks: a case study on sentiment analysis and fake review detection for marketing decision-making. Industrial Marketing Management, 90:523537, 2020.

JOSÉ RAMÓN TRILLO was born in Jaén (Spain) in 1994. He obtained his PhD in Intelligent Decision Support Systems from the University of Granada in 2024. He began his academic training in 2012 at the same university, where he simultaneously studied the degrees in Computer Engineering and Mathematics, completing both in 2018. After finishing his undergraduate studies, between 2018 and 2019, he pursued a Master's degree in Computer Engineering, delving into the advanced areas of artificial intelligence and information technologies. In 2022, he continued his education with a Master's degree in Mathematics, specialising in mathematical modelling applied to optimisation problems and computational algorithms. In 2024, he completed a Master's degree in Cybersecurity. Finally, in the same year, he completed his PhD at the University of Granada, contributing relevant research in the field of intelligent systems and automated decision-making.

MANUEL JESUS COBO received his M.Sc. and Ph.D. degrees in computer sciences from the University of Granada, Granada, Spain, in 2008 and 2011, respectively. He is an Associate Professor with the Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain. He developed the open-access software SciMAT for science mapping. His research interests include bibliometrics, science mapping analysis, social networks, artificial intelligence, data mining, and information science.

IGNACIO JAVIER PÉREZ GÁLVEZ was born in Granada, Spain, in 1984. He received his M.Sc. and Ph.D. degrees in computer science from the University of Granada, Spain, in 2007 and 2011, respectively. He is currently a Professor at the Department of Computer Sciences and Artificial Intelligence, University of Granada. His current research interests include group decision-making, consensus models, linguistic modelling, information aggregation, information retrieval, digital libraries, web quality evaluation, recommender systems, e-health, social media and Large Language Models. He is the Associate Editor of the Mathematics journal and author of over 30 papers published in international journals. Three of his papers are classified as highly cited and in the top 1% of the most cited papers in their field. His h-index is 18 according to the Web of Science, with more than 2.400 citations received.

JUAN ANTONIO MORENTE-MOLINERA was born in Granada, Spain, in 1986. He received the Ph.D. degree in information and communications technology from the University of Granada (UGR), Granada, Spain, in 2015. He is currently an Associate Professor with the Department of Computer Science and Artificial Intelligence, UGR. His H-index is 21 and he has published more than 40 papers in JCR journals. He is working in the decision making field trying to solve the challenges that the new era of ubiquitous computing brings to the decision support systems field. His research interests include group decision making, decision support systems, fuzzy ontologies, consensus models, linguistic modeling, and sentiment analysis. Dr. Morente-Molinera received the the UGR Best Thesis Prize in 2019. He is a member of the Soft Computing and Intelligent Information Systems Research Group.

FRANCISCO JAVIER CABRERIZO received his M.Sc. and Ph.D. degrees in computer science from the University of Granada, Spain, in 2006 and 2008, respectively. From 2008 to 2016, he was an Assistant Professor with the Department of Software Engineering and Computer Systems at the National Distance Education University, Madrid, Spain. He is currently a Full Professor with the Department of Computer Science and Artificial Intelligence at the University of Granada. His h-index is 34 and he presents more than 6 500 citations in Web of Science from Clarivate Analytics. His research interest includes fuzzy decision making, decision support systems, consensus models, linguistic modeling, aggregation of information, digital libraries, Web quality evaluation, and bibliometrics. Dr. Cabrerizo's awards and honors include the ISDA 2009 Conference Best Paper Award, the ICSEE 2015 Conference Best Paper Award, the 2014 Herbert Simon Award for outstanding contribution in International Journal of Information Technology & Decision Making, and the 2016 Andrew P. Sage Best Transactions Paper Award, which recognizes the authors of the best paper published annually in IEEE Transactions on Systems, Man, and Cybernetics: Systems. He has published extensively in leading international journals, and several of his papers are classified as Highly Cited Papers in the InCites Essential Science Indicators from Clarivate Analytics. He has been identified as Highly Cited Researcher by Clarivate Analytics from 2018 to 2022. He is an Associate Editor of IEEE Transactions on Cybernetics and International Journal of Information Technology & DecisionăMaking.

ENRIQUE HERRERA-VIEDMA (Fellow, IEEE) was born in 1966. He received the masters degree in mathematics, and the Ph.D. degree in mathematics, from the University of Granada, Granada, Spain, 1988 and 1988, respectively. He is a Full Professor of Computer Science and AI with the University of Granada, Granada, Spain, from 2007, where he is the Vice-President for Research and Knowledge Transfer from June 2015. He has published more than 560 peer-reviewed scientific publications, including 350 papers in JCR journals (More than 200 in Q1, more than 50 in D1). According to Web of Science, his H-index is 85 (more than 23 000 citations received) and according to Google Scholar his H-index is 101 (more than 41 000 citations received). He currently has 50 Highly Cited Papers indexed in Web of Science. Prof. Herrera-Viedma has been identified as Highly Cited Researcher in both categories, Computer Science and Engineering by Clarivate Analytics, in the years 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, and 2023. He received the UGR Young Researcher Career Award in 2006, the several UGR excellence research paper awards in 2003, 2010, 2019, and 2020, the 2011 IEEE Transactions on Fuzzy Systems Outstanding Paper Award from the IEEE Computational Intelligence Society, the 2016 Andrew P. Sage Best Transactions Paper Award, which recognizes authors of the best paper among all papers published in the IEEE Transactions on Systems, Man, and Cybernetics: Systems, and the 2014-Best Paper Herbert Simon Award of journal International Journal of Information Technology and Decision Making. He has been the Vice-President for Publications in the IEEE System, Man and Cybernetic Society from 2019 to 2020 and currently he is the New VP for Cybernetics till 2025. He is a Fellow of the Institute of Electrical and Electronic Engineers and also of International Fuzzy Systems Association for contributions to fuzzy decision making systems and linguistic modelling in 2021. He is also a member of the prestigious Spanish Think Tank Royal Institute Elcano supported by the large Spanish enterprises and the Spain Government.