electronics

MDPI

Article

Logic Operations for Assessment of Experts' Weight in Fuzzy Rule-Based Systems

Lydia Castronovo [1], Giuseppe Filippone [1], Giuseppe Giacopelli [1], Gianmarco La Rosa [1] and Marco Elio Tabacchi [1,2,*]

[1] Dipartimento di Matematica e Informatica, Università degli Studi di Palermo, Via Archirafi 34, 90123 Palermo, Italy; lydia.castronovo@unipa.it (L.C.); giuseppe.filippone01@unipa.it (G.F.); giuseppe.giacopelli@unipa.it (G.G.); gianmarco.larosa@unipa.it (G.L.R.)
[2] Istituto Nazionale di Ricerche Demopolis, 90123, Palermo, Italy , Italy
* Correspondence: marcoelio.tabacchi@unipa.it

Abstract

In Multi-Criteria Group Decision-Making (MCGDM), the assignment of weights to decision-makers is a crucial but methodologically delicate step, especially when the group includes both human experts and artificial experts such as intelligent agents, Artificial Intelligences (AIs) or Large Language Models (LLMs). Existing weighting strategies are often either difficult to interpret or poorly suited to heterogeneous groups of evaluators. In this paper, we investigate a fuzzy rule-based approach to expert weighting, building on a previously introduced methodological framework and focusing here on its application-oriented validation. The proposed method models expert weighting as a Fuzzy Rule-Based System (FRBS) in which the relevant properties of the experts are represented by linguistic variables and combined through interpretable IF–THEN rules. In this way, weighting policies can be expressed transparently and adapted to the requirements of the decision domain. The framework produces normalised weights in the interval $[0, 1]$, which can then be incorporated into standard MCGDM aggregation procedures. To assess the operational behaviour of the approach, we consider an application involving the weighting of four open-source LLMs (`apertus:8b`, `gemma4:e4b`, `mistral-small3.2:24b`, and `nemotron-cascade-2:30b`) over three multilingual criteria (English, Italian, Portuguese) and two resource-side criteria (VRAM, open-sourceness), each modelled by three trapezoidal fuzzy sets and combined into a five-class output partition; the underlying dataset is built from 10 independent repetitions of 100 questions per model. Under a language-focused rule base of five IF–THEN rules, the four experts receive sharply separated normalised weights $(0.003, 0.149, 0.301, 0.548)$—a top-to-bottom ratio above 180—whereas a combined linguistic/resource-aware rule base of five rules flattens the distribution to $(0.227, 0.360, 0.222, 0.191)$ and selects a different winner, demonstrating that policy changes are encoded explicitly in the output. A 100-run Kendall's τ perturbation analysis confirms that the induced rankings remain stable under moderate input noise, particularly for the language-focused policy, while substituting the Product t-norm with Gödel or Łukasiewicz leaves the language-focused ranking invariant but induces rank reversals in the more discriminative resource-aware policy. A comparison against three independent baselines (Markov Logic Networks, ProbLog, TOPSIS) shows that ProbLog reproduces the FRBS ordering in both case studies, MLN compresses the normalised scores under its global probabilistic interaction, and TOPSIS diverges whenever conditional IF–THEN preferences must be encoded. A worked end-to-end aggregation example with three alternatives, three criteria, and four experts further shows that the FRBS weights propagate into a clear selection of the best alternative, with aggregated scores $(S_1, S_2, S_3) = (8.88, 6.78, 6.21)$. These results confirm both the practical usability of the

Academic Editor: Ioannis Hatzilygeroudis

Received: 28 April 2026
Revised: 20 May 2026
Accepted: 30 May 2026
Published: 29 July 2026

method and its suitability for contexts in which multiple, potentially competing, objectives must be balanced explicitly. Overall, the paper provides an application-oriented study of an FRBS-based weighting scheme for artificial experts, highlighting its interpretability, adaptability, and potential relevance for contemporary MCGDM settings.

Keywords: Multi-Criteria Group Decision-Making; expert weighting; artificial experts; fuzzy rule-based systems; fuzzy inference; membership functions; interpretable decision support; large language models

1. Introduction

Multi-Criteria Group Decision-Making (MCGDM) extends Multi-Criteria Decision-Making (MCDM) to settings in which multiple decision-makers evaluate alternatives and a collective outcome is required [1]. In standard MCDM, criteria are identified and weighted according to their relative importance [2,3]. In MCGDM, an additional layer is needed, namely the weighting of the decision-makers themselves [1,4]. This issue is crucial because experts may differ in competence, objectives, and cognitive bias, and assigning equal importance to all evaluators may lead to distorted or unstable decisions [4–6].

The literature usually distinguishes between subjective weighting schemes, based on self-assessment or external judgment, and objective weighting schemes, based on mathematical indicators such as consistency, consensus, distance, or performance measures [3,7]. Despite their usefulness, both families present practical limitations, especially when weighting is perceived as a personal evaluation of the participating experts rather than as a modelling choice [1,5]. For this reason, recent research has explored hybrid and automated strategies in which artificial experts support or partially replace human experts in fuzzy MCGDM pipelines [8–10].

Artificial experts, such as optimisation algorithms, recommender systems, simulation-driven agents, and predictive models, are increasingly relevant in decision processes where scalability, repeatability, and traceability are required. In these settings, they should not be regarded merely as computational tools, but as decision contributors whose outputs may affect the collective ranking of alternatives. This motivates the need for principled weighting mechanisms capable of integrating human and artificial experts within the same aggregation pipeline, while preserving interpretability and operational flexibility. The broader interest in soft-computing approaches and their application to new domains is consistent with this perspective, as evidenced, among others, in [11,12].

In this work, an extension of the work in [10], we follow this direction and study a fuzzy framework in which expert weights are produced through a Fuzzy Rule-Based System (FRBS). The key idea is to encode weighting policies as interpretable IF–THEN rules. This allows us to obtain weights in the interval $[0, 1]$ that are transparent, adaptable, and suitable for mixed groups, including both human and artificial decision-makers.

A broader panoramic view places this proposal within the long development of fuzzy-based decision systems, from early rule-based approximate reasoning models and linguistic control architectures to more recent fuzzy MCDM and MCGDM procedures [13].

In this paper, we propose a new version of the state-of-the-art in fuzzy MCGDM [13]. This new method has been tested on a Large Language Model (LLM) scoring task. The FRBS has been able to weigh performances, technological requirements and the open-source nature of the LLM tested. In conclusion, it has been proven that the proposed method is flexible enough to address real-world tasks. Since the real world does not have a unique

direction of optimisation, in this contribution, it is proven that optimisation algorithms should not either.

The choice of an FRBS in this work is motivated by five specific properties that distinguish it from data-driven (entropy, deviation) and elicitation-based (AHP-like) weighting schemes:

1. Explicit policy encoding: the rule base is a direct, human-readable statement of the weighting policy; data-driven methods infer weights from statistical properties of the criterion matrix and cannot, by construction, encode policy preferences such as "open-source models should be favoured."
2. Interpretability and auditability: every weight produced by the FRBS is traceable back to a small set of activated rules with quantifiable firing strengths, which is increasingly important when AI agents act as decision contributors [8,14].
3. No training data required: unlike machine-learnt weighting schemes, the FRBS can be deployed in settings where past consensus decisions are not available.
4. Uniform treatment of heterogeneous experts: human and artificial experts are summarised by the same kind of antecedent tuple (Section 4.2), so they can be aggregated without introducing a separate machinery for each.
5. Modifiability: when policy priorities change (e.g., a regulatory shift on openness), the rule base is edited locally without retraining.

The cost of these properties is the manual authoring of the rule base, which we discuss at the end of Section 5.2 and in Section 6.

The remainder of this paper is organised as follows. Section 2 reviews the literature most relevant to expert weighting, fuzzy rule-based systems, and group decision-making. Section 3 introduces the mathematical background on fuzzy sets, FRBSs, and the MCGDM framework required for the subsequent developments. Section 4 presents the proposed methodology and describes the corresponding operational pipeline. Section 5 illustrates the approach through an application involving the ranking of four open-source LLMs. Finally, Section 6 summarises the main findings of the paper and outlines possible directions for future research.

2. Related Works

From the decision-analytic and data-driven viewpoint, user weighting is commonly addressed through TOPSIS [15], VIKOR [16–18], scoring systems, and train–test evaluation protocols. TOPSIS ranks users by closeness to an ideal profile, while VIKOR provides a compromise ranking between collective utility and individual regret. Scoring systems based on reinforcement learning and curriculum learning capture dynamic behaviour, long-term utility, and performance under increasing task difficulty; this line is consistent with performance-based expert weighting [19,20]. Train–test protocols improve robustness by separating calibration from out-of-sample evaluation, reducing overfitting and enabling fairer comparison across human experts and AI agents; related aggregation strategies include adaptive weighting and evaluator-reliability estimation [1,4].

A second relevant line of work concerns the role of Fuzzy Rule-Based Systems themselves. Classical contributions by Mamdani and by Takagi–Sugeno established FRBSs as interpretable mechanisms for modelling control and decision policies through linguistic IF–THEN rules [21–23]. Later studies clarified the semantics and design of fuzzy rules, and surveyed the evolution of FRBSs from control-oriented architectures to broader knowledge-based and decision-support settings [24–26]. More recent reviews highlight that FRBSs remain attractive when transparency, modularity, and explainability are important, especially in application domains where purely black-box optimisation is difficult to justify [27]. This makes FRBSs a natural candidate for expert-weighting problems, because they al-

low weighting criteria to be stated explicitly and revised without abandoning formal aggregation procedures.

Within MCGDM, the literature usually distinguishes subjective, objective, and hybrid approaches to weighting decision makers. Subjective approaches rely on direct judgements, pairwise comparisons, or elicited confidence assessments and are close in spirit to classical expert judgement and preference elicitation traditions [2,19,20]. Objective approaches infer weights from observable properties of the decision process, for example, consistency, consensus, dispersion, deviation, or reliability indicators [28,29]. Hybrid approaches combine both sources of information, often blending structural indicators with performance or consensus information in order to avoid the rigidity of purely subjective schemes and the impersonality of purely objective ones [4,7,30]. The survey literature confirms that no single family dominates across applications, and that weighting design remains strongly dependent on the decision context, the available data, and the desired level of interpretability [1,3,7].

This point becomes even more delicate when some evaluators are artificial experts rather than human experts. In decision-support systems, artificial experts may take the form of optimisation routines, learned predictors, recommender components, simulation agents, or explainable AI modules that generate scores, rankings, or recommendations [8,9,31]. Their inclusion raises specific methodological questions: unlike human experts, artificial experts can be highly repeatable and scalable, yet their outputs may depend on training data, design assumptions, or hidden model constraints; therefore, assigning them a weight is not merely a social choice, but also a modelling decision about trust, accountability, and comparability across heterogeneous decision sources [14,32]. For this reason, explainability is not an accessory feature but a central requirement when human and artificial expertise are aggregated in the same MCGDM pipeline [8].

A representative recent contribution along the cloud-model line is [33], which proposes a rough integrated asymmetric cloud model for large-group decision-making under multi-granularity linguistic preferences. Their setting differs from ours along three orthogonal axes.

1. Representation: they encode preferences as asymmetric clouds combined with rough approximations to capture randomness and the relevance of decision-makers, whereas we represent expert profiles as crisp inputs to a rule base over trapezoidal fuzzy partitions.
2. Weight determination: their decision-maker weights are computed objectively via a Shapley-function-based scheme over a trust-propagation graph, whereas our weights are produced by an interpretable rule base that explicitly encodes the policy of the decision-maker; the two schemes therefore solve different problems (objective dispersion-aware aggregation vs. policy-encoded weighting under interpretability constraints).
3. Group structure: their method targets large groups with mutually relevant evaluators, whereas we target small heterogeneous mixed groups, including artificial experts. The two approaches are thus complementary rather than substitutable, and a hybrid scheme combining the trust-propagation logic of [33] with the rule-based weighting proposed here is a natural direction for future work.

Against this background, the main gap addressed by the present study is not the absence of weighting methods per se, but the limited application-oriented validation of FRBS-based weighting schemes for artificial experts and the still scarce use of explicit sensitivity analyses for this specific setting. The methodological building blocks are available in the broader MCGDM and FRBS literature, and a first focused proposal for weighting artificial experts through fuzzy rules has already been presented in [10]; however, the literature still offers little evidence on how such a scheme behaves when transferred to realistic

decision tasks, compared with more conventional weighting strategies, and perturbed under different modelling assumptions. Since sensitivity analysis is widely recognised as a key instrument in MCDM and MCGDM for testing robustness, credibility, and dependence on parameter choices, this remains a non-trivial omission [34,35].

In this sense, the present article differs both from the previous methodological contribution and from the broader families of weighting methods already available in the literature. Relative to [10], the contribution here is steered less by the formal introduction of the FRBS mechanism and more by its validation in an application-oriented scenario inspired by the assessment of LLMs, with attention to how performance, technological requirements, and openness can be jointly encoded in the weighting process. Relative to standard subjective, objective, or hybrid schemes, the proposed strategy emphasises an interpretable rule base, making the rationale for assigning different weights to human and artificial experts more explicit and auditable. This orientation is consistent with broader attempts to move fuzzy methods toward new application domains while preserving semantic clarity and practical usability.

3. Preliminaries

In this section, the main preliminary notions concerning Fuzzy Set Theory (FST) and Fuzzy Rule-based Systems (FRBS) will be reviewed briefly. In the following section, the essential elements of the Multi-Criteria Decision-Making (MCDM) and Multi-Criteria Group Decision-Making (MCGDM) frameworks will be introduced. We refer the reader to Table 1 for the mathematical symbols used throughout this work.

Table 1. Table of mathematical symbols used throughout the paper.

Symbol	Meaning
U, U_X, U_Y	Universe of discourse (set on which a fuzzy set is defined)
$\mu_A : U \to [0,1]$	Membership function of fuzzy set A
$\mathcal{F}(U)$	Family of all fuzzy sets over U
$\otimes, \oplus$	Generic t-norm and t-conorm (Definitions 1 and 2)
$\Rightarrow, \ominus$	Generic fuzzy implication and negation
$\otimes_L, \otimes_G, \otimes_P, \otimes_Z$	t-norm of Łukasiewicz, Gödel, Product, Zadeh logic (Table 2)
R	Fuzzy relation; $\mu_R : U_X \times U_Y \to [0,1]$
$\widetilde{t} = (a,b,c)$	Triangular fuzzy number (Definition 3)
$\widetilde{t}_p = (a,b,c,d)$	Trapezoidal fuzzy number (Definition 4)
$A = \{A_1, \ldots, A_n\}$	Set of n decision alternatives
$C = \{C_1, \ldots, C_m\}$	Set of m criteria
a_{ij}	Score of alternative A_i on criterion C_j
$w_j \in [0,1]$	Weight of criterion C_j ($\sum_j w_j = 1$); index j ranges over criteria
x_i	Rank value of alternative A_i under WSM (Equation (11))
A^*	Optimal alternative (Equation (12))
$E_k, k = 1, \ldots, p$	The k-th expert (p experts total); index k ranges over experts
$\bar{\pi}_k$	Unnormalised expert weight: defuzzified output of the FRBS
$\pi_k \in [0,1]$	Normalised expert weight ($\sum_k \pi_k = 1$, Equation (14))
X_i	Aggregated score of A_i across experts, $X_i = \sum_k \pi_k x_i^{(k)}$ (Section 5.7)
$\bar{x}, \sigma$	Sample mean and standard deviation of a metric Section 4.2.1

Table 2. Table of t-norm, t-conorm, fuzzy implication and negation of Łukasiewicz, Gödel, Product, and Zadeh logics.

	Logic			
	Łukasiewicz	**Gödel**	**Product**	**Zadeh**
$x \otimes y$	$\max(x+y-1, 0)$	$\min(x, y)$	$x \cdot y$	$\min(x, y)$
$x \oplus y$	$\min(x+y, 1)$	$\max(x, y)$	$x+y-x \cdot y$	$\max(x, y)$
$x \implies y$	$\min(1-x+y, 1)$	$\begin{cases} 1 & \text{if } x \leq y, \\ y & \text{otherwise} \end{cases}$	$\min\left(1, \frac{y}{x}\right)$	$\max(1-x, y)$
$\ominus x$	$1-x$	$\begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{otherwise} \end{cases}$	$\begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{otherwise} \end{cases}$	$1-x$

3.1. Fuzzy Set Theory

In this section, we recall essential terminology from Fuzzy Set Theory (FST), which underpins FRBSs. For more details on fuzzy sets, we refer the reader to [36].

A fuzzy set [37] is specified by its membership function, namely a mapping from a Universe of Discourse (UoD) U to a complete lattice–typically the unit interval:

$$\mu_A \colon U \mapsto [0, 1], \tag{1}$$

where A denotes the (symbolic) label of the set and μ_A its (semantic) interpretation. In order to keep notation compact, one shall simply write "fuzzy set A" in place of "the fuzzy set labelled A with membership function μ_A," and one writes $\mu_A(x)$ for the membership degree of $x \in U$. The family of all fuzzy sets over U is denoted by $\mathcal{F}(U)$.

This intuitive approach can be extended [38] to a general Fuzzy set B endowed with a Fuzzy Membership function $\mu_B(x)$, through the formula:

$$\alpha = \max_{x \in U} \min\{\mu_A(x), \mu_B(x)\} \tag{2}$$

as can be easily seen in the case of a singleton set $B = \{x_0\}$:

$$\alpha = \max_{x \in U} \min\{\mu_A(x), \mu_B(x)\} = \max_{x \in U} \min\{\mu_A(x), \delta_{x_0}(x)\} = \mu_A(x_0) \tag{3}$$

where $\delta_{x_0}(x)$ is the discrete Dirac-delta membership, that is 1 evaluated in x_0 and zero elsewhere.

Fuzzy set connectives generalise intersection, union, and complement via t-norms (denoted as $\otimes$), t-conorms (s-norms) (denoted as $\oplus$), and negations (denoted as $\ominus$).

Definition 1 (cf. [39]). *A function* $\otimes \colon [0, 1] \times [0, 1] \to [0, 1]$ *is a* t-norm *if, for all* $x, y, z \in [0, 1]$,

$$y \leq z \Rightarrow x \otimes y \leq x \otimes z, \qquad \text{(Monotonicity)}$$
$$x \otimes (y \otimes z) = (x \otimes y) \otimes z, \qquad \text{(Associativity)}$$
$$x \otimes y = y \otimes x, \qquad \text{(Commutativity)}$$
$$x \otimes 1 = x. \qquad \text{(Boundary Condition)}$$

Definition 2 (cf. [39]). *A function* $\oplus\colon [0,1]\times[0,1]\to[0,1]$ *is a* t*-conorm if it satisfies the axioms of Definition 1 with the boundary condition replaced by*

$$x \oplus 0 = x, \qquad \text{(Boundary Condition)}$$

for all $x \in [0,1]$.

Three logics are commonly used: Łukasiewicz, Gödel, and Product Logic [40] (see Table 2). Zadeh's operators, defined by $a \otimes_Z b = \min(a,b)$ and $a \oplus_Z b = \max(a,b)$, coincide with Gödel logic for conjunction and disjunction, while differing in the treatment of negation and implication. In order to distinguish the logic used, we write the symbols in Table 3.

Table 3. Table of symbols used for t-norm, t-conorm, fuzzy implication and negation of Łukasiewicz, Gödel, Product, and Zadeh logics.

Logic			
Łukasiewicz	**Gödel**	**Product**	**Zadeh**
$\otimes_L$	$\otimes_G$	$\otimes_P$	$\otimes_Z$
$\oplus_L$	$\oplus_G$	$\oplus_P$	$\oplus_Z$
$\underset{L}{\Longrightarrow}$	$\underset{G}{\Longrightarrow}$	$\underset{P}{\Longrightarrow}$	$\underset{Z}{\Longrightarrow}$
$\ominus_L$	$\ominus_G$	$\ominus_P$	$\ominus_Z$

A *fuzzy relation* R on $U_X \times U_Y$ is a fuzzy subset of the Cartesian product:

$$\mu_R\colon U_X \times U_Y \mapsto [0,1]. \tag{4}$$

Beyond the usual fuzzy set operations, relations admit a *composition* operator, fundamental to fuzzy inference. Given R_1 on $U_X \times U_Y$ and R_2 on $U_Y \times U_Z$, their (sup–min) composition is

$$\mu_{R_2 \circ R_1}(x,z) = \sup_{y \in U_Y} \min\{\mu_{R_1}(x,y), \mu_{R_2}(y,z)\}. \tag{5}$$

Analogously, the composition of a fuzzy set A on U_X with a relation R on $U_X \times U_Y$ yields

$$\mu_{R \circ A}(y) = \sup_{x \in U_X} \min\{\mu_A(x), \mu_R(x,y)\}, \tag{6}$$

which is known as the *compositional rule of inference* (CRI) [41].

Intuitively, Equation (6) matches a fact (expressed as a fuzzy set A_X on U_X) with a relation R to infer a new fact (expressed as a fuzzy set A_Y on U_Y). Viewing the relation R as a fuzzy rule $A_X \to A_Y$, the fuzzy composition of relation underpins the *Generalised Modus Ponens* (GMP), which is the cornerstone of approximate reasoning:

$$\frac{A'_X, \quad A_X \to A_Y}{A'_Y} \tag{7}$$

which, for any premise $A'_X \in U_X$, infers a conclusion $A'_Y \in U_Y$.

Fuzzy Numbers

A fuzzy number is a specific type of fuzzy subset within the set of real numbers $\mathbb{R}$, characterised by additional defining properties (see, e.g., [40,42]). For these notions, we refer to the definitions and results used in the 1983 paper by P.J.M. van Laarhoven and W.

Pedrycz [43]. In this paper, we consider triangular and trapezoidal fuzzy numbers defined as follows.

Definition 3. *A* triangular fuzzy number (TFN) $\widetilde{t} = (a, b, c)$, *with* $a, b, c \in \mathbb{R}$ *and* $a \leq b \leq c$, *is a fuzzy number whose membership function has a trapezoidal shape, i.e., for every* $x \in \mathbb{R}$,

$$\widetilde{t}(x) = (a, b, c)(x) = \mu(x) = \begin{cases} \frac{x-a}{b-a}, & x \in [a, b], \\ 1 & x = b, \\ \frac{x-b}{c-b}, & x \in [b, c], \\ 0 & \textit{otherwise}. \end{cases} \tag{8}$$

Definition 4 (cf. [44], Definition 2.3)**.** *A* trapezoidal fuzzy number *(TpFN)* $\widetilde{t_p} = (a, b, c, d)$, *with* $a, b, c, d \in \mathbb{R}$ *and* $a \leq b \leq c \leq d$, *is a fuzzy number whose membership function has a trapezoidal shape, i.e., for every* $x \in \mathbb{R}$,

$$\widetilde{t_p}(x) = (a, b, c, d)(x) = \mu(x) = \begin{cases} \frac{x-a}{b-a}, & x \in [a, b], \\ 1 & x \in [b, c], \\ \frac{x-d}{c-d}, & x \in [c, d], \\ 0 & \textit{otherwise}. \end{cases} \tag{9}$$

3.2. Fuzzy Rule-Based Systems

Fuzzy sets have been used in many applications, but their most notable success is undoubtedly in control systems. Many control systems employ the well-known Fuzzy Rule-Based Systems (FRBSs), which exploit a rule base formed by a simple set of IF–THEN rules in the form

$$\text{IF } \texttt{antecedents} \text{ THEN } \texttt{consequent}, \tag{10}$$

where the `antecedents` are usually a composition of logical operations (`AND`, `OR`, and `NOT`) of fuzzy sets, while the `consequent` is usually a fuzzy set. This particular setting of `antecedents` and `consequent` is typical of Mamdani FRBSs [21]. For more details on FRBSs, we refer the reader to [26,27].

In particular, FRBSs are explainable systems composed of (see Figure 1 for a graphical representation):

- **Fuzzifier**: The fuzzifier receives the real-world input to the fuzzy system. In the fuzzy-systems literature, this is commonly termed a *crisp* input because it represents an exact value of the parameter under consideration. The role of the fuzzifier is to map this precise quantity onto linguistic categories, such as “large”, “medium”, or “high”, by assigning an appropriate degree of membership. This degree typically lies within the real interval $[0, 1]$.
- **Knowledge Base**: The Knowledge Base constitutes the core of the fuzzy system and comprises both the database and the rule base. The database specifies the membership functions associated with the fuzzy sets employed in the rules, whereas the rule base contains a collection of fuzzy IF–THEN statements.
- **Inference System**: The Inference System, also referred to as the decision-making unit, carries out the reasoning process over the rule set. Its function is to determine how the rules are evaluated and combined in order to generate the fuzzy output.
- **Defuzzifier**: The output produced by the inference stage is inherently fuzzy. Since practical applications generally require a *precise* output, the defuzzifier converts this fuzzy result into a crisp value that can be interpreted in real-world terms. In this respect, it performs the inverse operation of the input stage.

In essence, FRBSs produce mathematical functions in multiple variables to model almost any complex control system by employing this simple rule base. For instance, if each rule in the rule base contains two `antecedents` and one `consequent`, it defines a three-dimensional surface over the fuzzy-set domains associated with the antecedent and consequent variables.

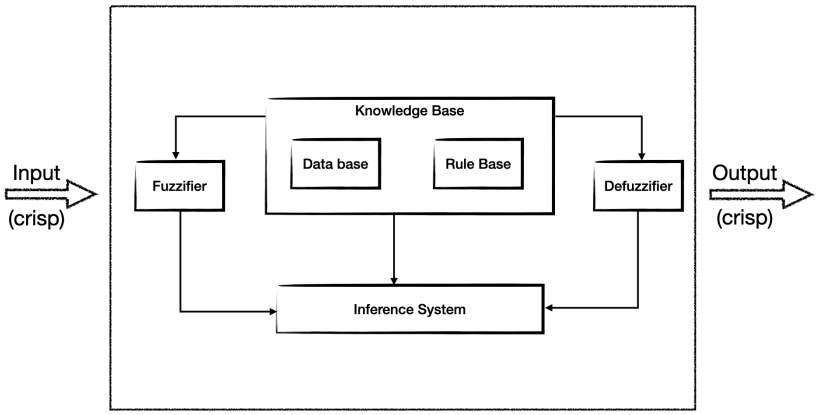

Figure 1. Main components of a Fuzzy Rule-Based System: fuzzification of crisp inputs, storage of rules and membership functions in the Knowledge Base, rule processing by the Inference System, and conversion of fuzzy outputs into crisp values through defuzzification.

3.3. Multi-Criteria Group Decision-Making

In this section, we briefly recall the basic setting of Multi-Criteria Decision-Making (MCDM), i.e., a branch of decision theory concerned with the comparison of several feasible alternatives in order to identify the optimal one, namely the option that best satisfies given goals, objectives, values, and priorities.

In formal terms, one considers a collection of *n alternatives* $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ and a family of *m criteria* $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$. The latter, sometimes referred to as attributes or goals, specify the dimensions under which the available alternatives are evaluated.

For each pair (A_i, C_j), a value – denoted as score – a_{ij} is assigned in order to express the degree to which the alternative A_i satisfies the criterion C_j.

The purpose of this framework is to determine the best alternative, denoted as A^*. In order to achieve this objective, each criterion C_j is also associated with a coefficient $w_j \in [0, 1]$, representing its importance in the decision problem. These coefficients are commonly normalised in such a way that $\sum_{j=1}^{m} w_j = 1$. Their determination may depend either on the judgement of a single decision-maker or on an aggregated assessment provided by several experts. Methods for deriving and eliciting such weights have been extensively studied in the literature; see, for instance, [45].

In addition to identifying the optimal alternative, MCDM typically yields an overall value x_i, usually called *rank*, for each alternative A_i, with $i = 1, \ldots, n$. This value summarises the global goodness of the alternative and makes it possible to rank all feasible options according to their overall suitability within the decision problem.

The information concerning alternatives and criteria is often arranged in a decision matrix, where rows are indexed by the alternatives and columns by the criteria. Such a matrix compactly collects the entries (scores) a_{ij} and provides the starting point for the application of aggregation procedures, combining criterion evaluations with their corresponding weights.

Among the classical techniques employed to determine this ranking value, one may consider the simplest one, namely the Weighted Sum Method (WSM). In this case, for every $i = 1, \ldots, n$, the score assigned to A_i is given by

$$x_i = \sum_{j=1}^{m} a_{ij} w_j. \tag{11}$$

For further details on alternative methods used to determine ranking values, we refer the reader to studies based on triangular aggregation operators [46], Simple Additive Weighting (SAW) [47,48], TOPSIS [49], VIKOR and TODIM [50,51], Analytic Hierarchy Process (AHP) and fuzzy-integral approaches [52,53], ELECTRE [54,55], and PROMETHEE [56,57].

The alternatives are then ranked from the largest to the smallest value of x_i, and the preferred choice is the one corresponding to the maximum score, that is,

$$A^* = \arg\max_{A_i} x_i. \tag{12}$$

When multiple experts are involved, the setting is referred to as Multi-Criteria Group Decision-Making (MCGDM), in which the optimal alternatives identified by the individual experts must be combined into a single overall optimal alternative. Let p denote the number of experts, and let E_k, for $k = 1, \ldots, p$, be the l-th expert. Each expert E_k determines an optimal alternative A_k^* by means of an MCDM method. The method adopted by one expert need not coincide with those adopted by the others, so that different experts may employ different MCDM procedures.

Accordingly, the aggregator (see, e.g., [44]) is a function that associates the p optimal alternatives $A_1^*, \ldots, A_p^*$ with a new alternative, denoted by A^* for the reader's convenience, representing the overall optimal choice. In order to determine this aggregated optimal alternative, a weight π_k is assigned to each expert E_k, for $k = 1, \ldots, p$. These weights are usually normalised so that $\sum_{k=1}^{p} \pi_k = 1$, and act as scalar coefficients expressing the relative importance attributed to the optimal alternative A_k^* selected by the expert E_k. Thus, π_k may be interpreted as a measure of the expertise of E_k, so that its optimal alternative has a greater influence on the determination of the aggregated optimum A^*.

In general, no universal constraints are imposed on the form of the aggregator, as its definition depends on the specific decision problem under consideration. For example, if the problem concerns the location of a nuclear power plant and the alternatives correspond to physical sites, then the aggregator cannot reasonably be defined as a "mean"—with weights $\{\pi_1, \ldots, \pi_p\}$—of the alternatives A_k^*, since the resulting alternative A^* might fail to belong to the set $\{A_1^*, \ldots, A_p^*\}$. In such a case, a more appropriate aggregator could be one that minimises the distance between the aggregated optimum A^* and the set of alternatives $\{A_1^*, \ldots, A_p^*\}$.

If the mean is adopted as the aggregation operator, then the final alternative is given by

$$A^* = \sum_{k=1}^{p} \pi_k A_k^*. \tag{13}$$

Finally, it is possible to extend the MCDM and MCGDM using fuzzy logic. For instance, fuzzy sets can be employed for the scores a_{ij} of the decision matrix and/or the weights w_j of the criteria. For MCGDM, the weights π_k of the experts can also be handled with this methodology. In this case, the frameworks are referred to as Fuzzy MCDM and Fuzzy MCGDM, respectively.

4. Proposed Method

In this section, we present our method for determining, in an MCGDM framework with p experts, the weights π_k associated with the experts E_k, for $k = 1, \ldots, p$.

First, we provide a brief overview of the method proposed in [10].

4.1. Background: Preliminary FRBS Weighting Framework

Firstly, a concise overview of the methodology employed in [10] will be provided. The referenced study proposed a preliminary framework for the weighting of artificial experts in the context of MCGDM or Fuzzy MCGDM, utilising a Larsen-like FRBS architecture. The fundamental proposition underlying this study was that, in scenarios where artificial users are utilised in lieu of or in conjunction with human experts, their contributions should not be aggregated indiscriminately. Instead, it is recommended that each expert (artificial and human) be assigned a weight that reflects properties relevant to its anticipated behaviour within the decision process. This perspective was motivated by the observation that no ranking method is uniformly preferable across all decision settings and that the performance of a given method may vary substantially according to the structure of the problem, the criteria considered, and the stability of the rankings it produces [45].

Within that framework, the weighting problem was reformulated as a fuzzy inference task. Artificial experts were described through a collection of linguistic variables representing salient characteristics, e.g., their precision. These properties were modelled by means of triangular membership functions and combined through an interpretable rule base of IF–THEN rules in order to derive a crisp (or fuzzy) weight π_k for each expert. The resulting output could then be defuzzified and normalised, thus making it directly usable within standard MCGDM procedures, while still preserving the possibility of a fully fuzzy treatment whenever required by the application, e.g., in [44,58].

Consequently, the preceding contribution established a structured and extensible basis for the evaluation of artificial experts. The primary value of the study was not limited to the specific example examined. It demonstrated that the allocation of expert weights can be facilitated through a transparent, rule-based mechanism that is capable of incorporating heterogeneous criteria and qualitative assessments. In particular, the selection of linguistic variables and their fuzzy sets exhibited consistency with the artificial experts that were taken into account. Concurrently, the framework was deliberately proffered as an inaugural methodological step, thereby leaving the question of how such a weighting scheme might be specialised, refined, or operationalised in more targeted decision contexts unresolved.

The present study builds on this foundation, progressing from the general problem of weighting artificial experts to the development of a more focused methodological contribution. In this sense, the earlier framework should be understood as the conceptual and inferential substrate upon which the current proposal is constructed. It provides the rationale for differentiating among artificial decision-makers, the formal language through which such differentiation can be expressed, and the methodological bridge between qualitative expert descriptors and quantitatively deployable weights in subsequent group decision-making models.

4.2. Operational Pipeline of the Extended Method

In this section, we describe the extension of the procedure adopted in [10] to produce either a crisp or a fuzzy weight for a given decision domain. We refer the reader to Algorithm 1 for a schematic presentation of the proposed method, and to Figure 2 for a corresponding flowchart of the overall pipeline.

Algorithm 1 FRBS-based expert weighting (extended methodology, with optional MCGDM aggregation)

Require: Set of experts $\mathcal{E} = \{E_1, \ldots, E_p\}$, each with a crisp profile $\left(x_1^{(k)}, \ldots, x_l^{(k)}\right)$. *(Optional)* decision matrix $X = [x_{ij}^{(k)}] \in \mathbb{R}^{m \times n \times p}$ over m alternatives and n criteria, and criterion weights $v_1, \ldots, v_n \in [0,1]$ with $\sum_j v_j = 1$.

Ensure: Normalised expert weights $\pi_1, \ldots, \pi_p \in [0,1]$ with $\sum_k \pi_k = 1$. *(Optional)* aggregated scores $S_1, \ldots, S_m$ and best alternative A^*.

Phase 1—Domain modelling (Section 4.2).

Select the antecedent variables $(X_1, \ldots, X_l)$ describing the experts; for mixed groups including human experts, follow Section 4.2.2 (e.g., `Years of Experience`, `Domain Expertise`, `Past Consistency`, `Self Confidence`).

For each antecedent X_i, construct the trapezoidal partition $\{\texttt{Low}_i, \texttt{Medium}_i, \texttt{High}_i\}$.

Fix the consequent variable `Weight` on $[0, 10]$, partitioned into the five fuzzy sets {`Very Low`, `Low`, `Medium`, `High`, `Very High`}.

Phase 2—Rule-base authoring.

Encode the weighting policy as a finite rule base $\mathcal{R} = \{r_1, \ldots, r_q\}$ of IF–THEN rules. Combine antecedents through a t-norm $\otimes$ (AND) and a t-conorm $\oplus$ (OR).

Phase 3—Inference and defuzzification.

for each expert $E_k \in \mathcal{E}$ **do**

  Fuzzify the crisp profile $\left(x_1^{(k)}, \ldots, x_l^{(k)}\right)$ to obtain membership degrees.

  for each rule $r_i \in \mathcal{R}$ **do**

    Compute the firing strength $\alpha_{ik} \in [0,1]$ from antecedent memberships via the chosen $\otimes, \oplus$.

    Apply the chosen implication to obtain the rule consequent.

  end for

  Aggregate all consequent fuzzy sets and defuzzify via the FAME-like centre-of-gravity [59]:

$$\bar{\pi}_k = \frac{\sum_i \alpha_{ik} \int y \, w_i(y) \, \mathrm{d}y}{\sum_i \int w_i(y) \, \mathrm{d}y},$$

  where $w_i(y)$ are the trapezoidal consequent membership functions.

end for

Phase 4—Normalisation.

Normalise the crisp weights via Equation (14): $\pi_k = \bar{\pi}_k / \sum_{i=1}^{p} \bar{\pi}_i$ for $k = 1, \ldots, p$.

Phase 5—(Optional) MCGDM aggregation (Section 5.7).

if a decision matrix X and criterion weights v_j are provided **then**

  Compute the group decision matrix $\bar{x}_{ij} = \sum_{k=1}^{p} \pi_k \, x_{ij}^{(k)}$ (Equation (29)).

  Compute alternative scores $S_i = \sum_{j=1}^{n} v_j \, \bar{x}_{ij}$ (Equation (30)).

  Select $A^* = \arg\max_i S_i$ (Equation (31)).

end if

return $\pi_1, \ldots, \pi_p$ (and $S_1, \ldots, S_m, A^*$ if Phase 5 was executed).

Crucially, the linguistic variables and rule base are designed *independently for each application domain*, reflecting the relevant properties of the experts to be weighted.

For each domain under consideration, the method follows a structured pipeline composed of four main phases.

First, we define the domain-specific linguistic variables. Let $(X_1, \ldots, X_l)$ denote the l antecedent variables selected to characterise the experts in the considered domain. Each variable is defined in a universe of discourse that is appropriate for the application context and is partitioned into fuzzy sets, so as to capture the linguistic assessments relevant to that domain. In addition, we introduce the consequent variable, denoted by `Weight`, whose universe is fixed to the interval $[0, 10]$.

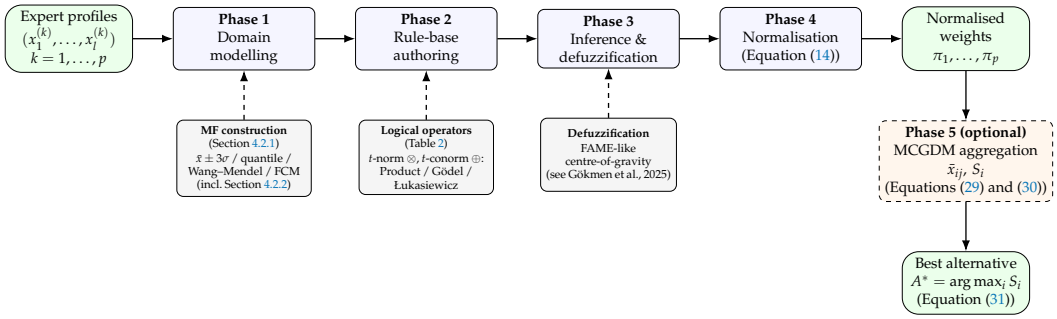

Figure 2. Workflow of the extended FRBS expert-weighting framework. The solid horizontal pipeline traces the four main phases of the methodology, from the expert profiles $(x_1^{(k)}, \ldots, x_l^{(k)})$ to the normalised weights π_k. The grey side blocks summarise the modelling choices available at each stage: trapezoidal-MF construction (Gaussian-anchored, quantile-based, Wang–Mendel histogram, or FCM clustering), the t-norm/t-conorm pair, and the FAME-like centre-of-gravity defuzzifier. The dashed-border block (Phase 5) shows the optional Weighted-Sum-Method MCGDM aggregation, which propagates the FRBS weights into the final alternative score S_i and selects the best alternative A^*. For the Defuzzification phase, we refer the reader to [59].

Second, we formulate the domain-specific fuzzy IF–THEN rules. These rules encode the relationship between the antecedent variables and the resulting expert weight. The antecedent part of each rule may combine linguistic conditions through the logical connectives `AND` and `OR`. In particular, conjunction is modelled by means of the t-norm $\otimes$, whereas disjunction is represented through some t-conorm $\oplus$, as those given in Table 2. The consequent part of each rule assigns the output to an appropriate `Weight` class.

Third, once the rule base has been defined, we perform inference and defuzzification. For each expert E_k, described by the crisp profile $(x_1^{(k)}, \ldots, x_l^{(k)})$, where $x_i^k \in U(X_i)$, i.e., x_i^k belongs to the Universe of Discourse of X_i, the input values are first fuzzified with respect to the previously defined antecedent fuzzy partitions. The firing strength of each rule is then computed according to the membership degrees of the inputs and the adopted fuzzy connectives (`AND`/`OR`). Next, the implication is applied to the consequent fuzzy sets, obtaining the corresponding firing strengths. The resulting outputs of all activated rules are subsequently aggregated. Finally, for each expert, a crisp weight $\bar{\pi}_k$ is obtained through defuzzification of the `Weight` through a CoG-like approach that uses the most recent advances in Fuzzy Additive Models [59]. In detail, the defuzzification used is:

$$\bar{\pi}_k = \frac{\sum_i \alpha_{ik} \int y w_i(y) dy}{\sum_i \int w_i(y) dy}$$

where $\alpha_{ik} \in [0, 1]$ is the firing strength of each rule i for each expert k, and $w_i(y)$ are the trapezoidal membership functions of the consequent.

Fourth, the obtained weights are normalised in order to derive relative expert weights $(\pi_1, \ldots, \pi_k)$. Specifically, for each expert E_k, the normalised weight is computed as:

$$\pi_k = \frac{\bar{\pi}_k}{\sum_{i=1}^{p} \bar{\pi}_i}. \tag{14}$$

This final step ensures that the resulting weights are expressed on a comparable scale and can be directly interpreted as relative contributions within the set of experts.

4.2.1. Construction of Antecedent Membership Functions from Empirical Data

In this section, we present the procedure adopted to model the antecedent variables by means of triangular and trapezoidal fuzzy numbers.

More specifically, the antecedent variables are derived from empirical results associated with a given metric M, such as *precision@10*. In this case, all results are available, and it is therefore possible to compute, for each metric under consideration, the pair $(\bar{x}, \sigma)$, where $\bar{x}$ denotes the mean and σ the standard deviation. Assuming that the metric values are approximately Gaussian, a natural mapping from the metric M to a corresponding fuzzy number.

In a three-partition fuzzy-set scheme, that is, whenever a linguistic variable is described by the standard family of three fuzzy sets `Low`, `Medium`, and `High`, let M be a metric and let $\mathbf{x} = (x_1, \ldots, x_n)$ denote the vector of observed values of M over n rounds. Let $\bar{x}$ be the mean of $\mathbf{x}$, and let σ be its standard deviation. Moreover, let $\mathcal{N}(\bar{x}, \sigma^2)$ denote the Gaussian distribution with mean $\bar{x}$ and standard deviation σ. Then, a suitable approximation of `Low`, `Medium`, and `High` by means of triangular fuzzy numbers is

$$\texttt{Low} = \tilde{t}(\bar{x} - 3\sigma, \bar{x} - 3\sigma, \bar{x}), \tag{15}$$

$$\texttt{Medium} = \tilde{t}(\bar{x} - 3\sigma, \bar{x}, \bar{x} + 3\sigma), \tag{16}$$

$$\texttt{High} = \tilde{t}(\bar{x}, \bar{x} + 3\sigma, \bar{x} + 3\sigma). \tag{17}$$

A similar approximation can be defined using trapezoidal fuzzy numbers:

$$\texttt{Low} = \widetilde{t_p}\left(\bar{x} - 3\sigma, \bar{x} - 3\sigma, \bar{x} - 2\sigma, \bar{x} - \frac{\sigma}{2}\right), \tag{18}$$

$$\texttt{Medium} = \widetilde{t_p}\left(\bar{x} - 2\sigma, \bar{x} - \frac{\sigma}{2}, \bar{x} + \frac{\sigma}{2}, \bar{x} + 2\sigma\right), \tag{19}$$

$$\texttt{High} = \widetilde{t_p}\left(\bar{x} + \frac{\sigma}{2}, \bar{x} + 2\sigma, \bar{x} + 3\sigma, \bar{x} + 3\sigma\right). \tag{20}$$

It is worth noting that the above partitions define a Ruspini [60] partition such that

$$\sum_{x \in \mathbf{x}} \mu_{\texttt{Low}}(x) + \mu_{\texttt{Medium}}(x) + \mu_{\texttt{High}}(x) = 1. \tag{21}$$

In particular, the universe of discourse of both fuzzy representations is defined over the interval from $\bar{x} - 3\sigma$ to $\bar{x} + 3\sigma$, which corresponds to approximately 99.7% of the area under $\mathcal{N}(\bar{x}, \sigma^2)$. For trapezoidal fuzzy numbers, the plateau is taken as the interval from $\bar{x} - \frac{\sigma}{2}$ to $\bar{x} + \frac{\sigma}{2}$, which corresponds to approximately the 38.3% of the area under $\mathcal{N}(\bar{x}, \sigma^2)$. This choice yields a simple and interpretable fuzzy partition of the observed range of the metric into three overlapping linguistic regions. In Figure 3, we provide a graphical representation of both approximations.

The $\bar{x} \pm 3\sigma$ construction above is a support-anchoring heuristic and does not assume that the underlying distribution is Gaussian: the Ruspini partition property holds for any trapezoidal supports satisfying the boundary conditions. Nevertheless, when the sample is small ($n = 10$ in our experiments) or visibly skewed or multimodal, both $\bar{x}$ and σ become noisy estimates, and the resulting partitions inherit this uncertainty. Several distribution-free alternatives are available and produce equivalent partitions when the data are approximately symmetric and unimodal:

- **Quantile-based supports:** replacing $(\bar{x} - 3\sigma, \bar{x} - 2\sigma, \bar{x} - \sigma/2, \bar{x} + \sigma/2, \bar{x} + 2\sigma, \bar{x} + 3\sigma)$ with the empirical quantiles $(q_{0.5\%}, q_{5\%}, q_{30\%}, q_{70\%}, q_{95\%}, q_{99.5\%})$, which are robust to skew and outliers.
- **Histogram-based partitioning** along the lines of Wang–Mendel [61], in which fuzzy-set centroids are placed at modes of an empirical histogram.
- **Clustering-based partitioning** via Fuzzy C-Means (FCM) [62,63] with $C = 3$ on the observed metric values, detailed below.

In particular, in clustering-based partitioning, given the observed values $\mathbf{x} = (x_1, \ldots, x_n)$ of metric M, FCM minimises the objective function J given as

$$J = \sum_{i=1}^{n} \sum_{j=1}^{C} u_{ij}^2 \|x_i - v_j\|^2, \tag{22}$$

with

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}}} \in [0, 1], \tag{23}$$

where $m \in (1, \infty)$ is the hyper-parameter controlling the degree of fuzziness of the resulting clusters; in the FCM setting, it is commonly chosen as $m = 2$. FCM returns three centroids $\{v_1, v_2, v_3\}$ that we sort in ascending order, i.e., $v_1 \leq v_2 \leq v_3$, and relabel as the prototypes of `Low`, `Medium`, `High`, respectively. Let $\mu_{12} = (v_1 + v_2)/2$ and $\mu_{23} = (v_2 + v_3)/2$ denote the midpoints between adjacent centroids, and let $x_{\min}, x_{\max}$ be the observed extrema. The trapezoidal fuzzy numbers are then defined as

$$\texttt{Low} = \widetilde{t}_p\left(x_{\min}, x_{\min}, v_1, \tfrac{v_1+\mu_{12}}{2}\right), \tag{24}$$

$$\texttt{Medium} = \widetilde{t}_p\left(\tfrac{v_2+\mu_{12}}{2}, v_2 - \tfrac{v_2-\mu_{12}}{2}, v_2 + \tfrac{\mu_{23}-v_2}{2}, \tfrac{v_2+\mu_{23}}{2}\right), \tag{25}$$

$$\texttt{High} = \widetilde{t}_p\left(\tfrac{v_3+\mu_{23}}{2}, v_3, x_{\max}, x_{\max}\right), \tag{26}$$

i.e., the plateau of each fuzzy set is centred on its cluster centroid with half-width set to half the distance to the nearest adjacent midpoint, and the outer supports are anchored at the observed extrema. The midpoints μ_{12}, μ_{23} are the crossover points at which consecutive fuzzy sets meet at membership 0.5. The resulting partition is strictly data-driven: no assumption is made on the shape of the underlying distribution, and the trapezoid widths automatically reflect the empirical cluster separations.

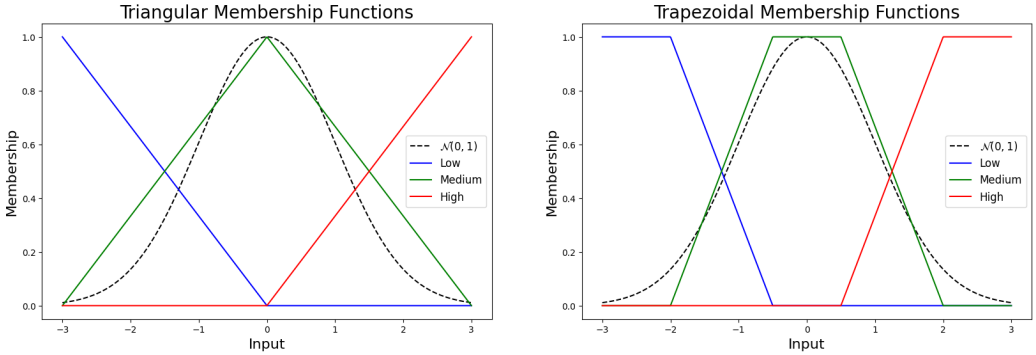

Figure 3. Three-partition fuzzy-set approximating the Gaussian distribution function $\mathcal{N}(0, 1)$ as triangular (on the **left**) and trapezoidal (on the **right**) fuzzy numbers `Low`, `Medium`, and `High`.

4.2.2. Mixed Groups Including Human Experts

The pipeline of Section 4.2 does not assume that the antecedent variables refer to artificial experts. For human experts, suitable linguistic variables include `Years of Experience`, `Domain Expertise` (e.g., from peer assessment or self-rating on a Likert scale), `Past Consistency` (e.g., Kendall τ between an expert's past rankings and the consensus on resolved cases), and `Self Confidence` on the current decision. Each variable is partitioned into trapezoidal fuzzy sets following Section 4.2.1 (using either the Gaussian-anchored heuristic or one of the data-driven alternatives discussed there). The rule base is then

expressed in the same IF–THEN syntax. In a mixed group, the same FRBS can include rules referring jointly to artificial and human attributes, e.g.,

$$\texttt{IF Domain Expertise is High AND Open-Sourceness is Not Applicable THEN Weight is High}, \quad (27)$$

provided that `Not Applicable` values are handled either by per-rule masking or by introducing a dedicated `Not Applicable` fuzzy set with degenerate membership.

5. Applications

In this section, we present our experiments using a FRBS for weighting four LLMs:

- `apertus:8b` [64],
- `gemma4:e4b` [65],
- `mistral-small3.2:24b` [66],
- `nemotron-cascade-2:30b` [67].

5.1. Dataset Introduction

The dataset has been created starting from the SDF dataset. Each LLM has been asked to answer 100 questions, and for every LLM, this process has been repeated 10 times. From these repetitions, the quantities $\bar{x}$ and σ have been calculated for each language, where $\bar{x}$ denotes the average number of correct answers and σ denotes the corresponding standard deviation. In this task, allowing a tolerance of $\pm 10\%$ is reasonable because the answers are numerical and small deviations from the target value should not be treated as fully incorrect. For this reason, the dataset used in the following examples is built from the measurements obtained under this tolerance level. These values are later transformed into trapezoidal fuzzy numbers (as defined in Section 4.2.1).

In Table 4 the multilingual results as $\bar{x} \pm \sigma$ are shown, where $\bar{x}$ is the average number of correct answers among the 10 runs, and σ is the standard deviation of these runs. In Table 5, we provide the resource-side criteria, namely the `VRAM` requirement and the qualitative open-source score adopted in the methodology.

Table 4. Multilingual results for `English`, `Italian`, and `Portuguese` as $\bar{x} \pm \sigma$, where $\bar{x}$ is the average number of correct answers among the 10 runs, and σ is the standard deviation of these runs.

Model	English ($\bar{x} \pm \sigma$)	Italian ($\bar{x} \pm \sigma$)	Portuguese ($\bar{x} \pm \sigma$)
`apertus:8b`	10.8 ± 2.440	10.1 ± 3.348	9.2 ± 1.619
`gemma4:e4b`	16.8 ± 1.317	13.4 ± 2.716	14.7 ± 2.669
`mistral-small3.2:24b`	18.4 ± 3.340	16.7 ± 3.401	15.6 ± 3.062
`nemotron-cascade-2:30b`	22.2 ± 2.348	16.4 ± 3.204	20.4 ± 2.591

Table 5. Resource-side criteria used in the examples.

Model	VRAM (kMiB)	Open-Sourceness
`apertus:8b`	9.5	10.0
`gemma4:e4b`	11.2	7.5
`mistral-small3.2:24b`	20.6	5.0
`nemotron-cascade-2:30b`	25.0	7.5

The dataset already suggests the qualitative trend later captured by the fuzzy method (see Figures 4 and 5): `nemotron-cascade-2:30b` is strongest on multilingual quality, `mistral-small3.2:24b` remains competitive on the language dimensions, while `apertus:8b` and `gemma4:e4b` are more favourable on the resource side.

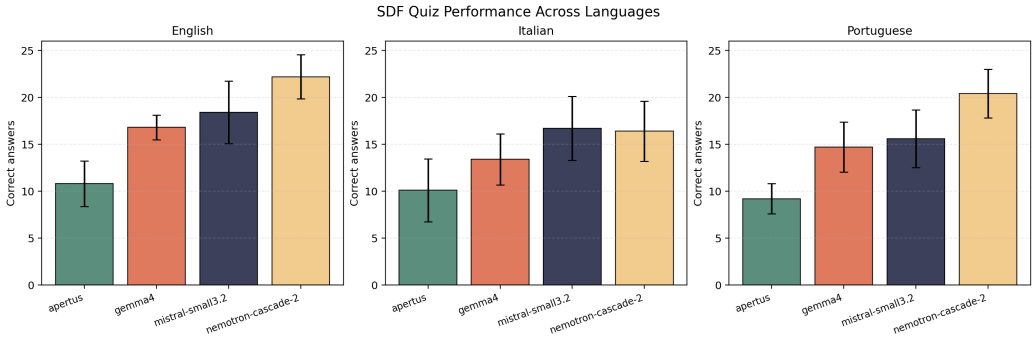

Figure 4. Language-side dataset reconstructed from the experimental summaries. In each barplot, the bar height represents the mean $\bar{x}$ and the whiskers represent the standard deviation σ.

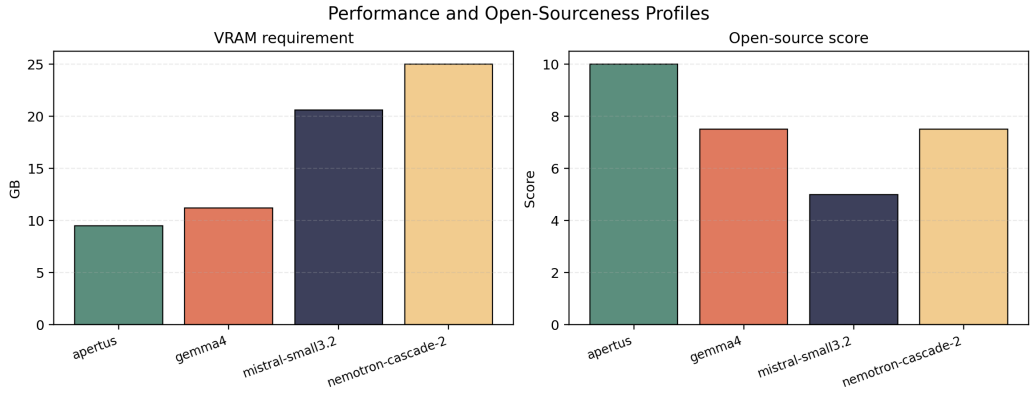

Figure 5. Resource-side criteria used together with the multilingual scores.

The `VRAM` score was assigned on the basis of the VRAM allocation size, expressed in kMiB, on a GPU equipped with 32 GB of VRAM. It was verified that RAM offloading was negligible and therefore did not materially affect the assessment.

By contrast, the `Open-Sourceness` score was determined from the following considerations. The model `apertus:8b` is both open-weight and open-source, in the sense that the entire training pipeline is openly available. For this reason, it was assigned a score of 10. The LLMs `nemotron-cascade-2:30b` and `gemma4:e4b` provide open weights and are available for commercial use; accordingly, each was assigned a score of 7.5. Moreover, these models have consistently remained available for commercial use over time.

The case of `mistral-small3.2:24b` is more nuanced. Although the model was released under an open-weight licence, specifically Apache 2.0, its commercial use was initially subject to restrictions. While its weights have only recently become available for commercial use, this evolving licensing regime may constitute a significant source of uncertainty for anyone intending to build a business upon this LLM. For this reason, it was assigned a score of 5.

5.2. Fuzzy Inference System

The fuzzy inference scheme takes five linguistic variables into consideration for the input and one for the output. In particular, the input variables are `English`, `Italian`, `Portuguese`, `VRAM`, and `Open-Sourceness`, while the output variable is `Weight`. The inputs `VRAM` and `Open-Sourceness` are resource-side criteria.

For each input variable, we defined three trapezoidal fuzzy numbers `Low`, `Medium`, and `High`. In particular, the linguistic input variables `English`, `Italian`, and `Portuguese` are represented with the trapezoidal parameters given Table 6 (see Figure 6 for a graphical representation). On the other hand, the resource-side criteria `VRAM` and `Open-Sourceness`

are represented using the trapezoidal parameters given in Table 7 (see Figure 7 for a graphical representation).

Table 6. Linguistic criteria membership function parameters. Each row corresponds to a trapezoidal fuzzy number and its parameters.

Linguistic Variable	Fuzzy Set	Parameters
English	Low	(2.823, 2.823, 7.565, 14.679)
	Medium	(7.565, 14.679, 19.421, 26.535)
	High	(19.421, 26.535, 31.277, 31.277)
Italian	Low	(4.899, 4.899, 7.982, 12.608)
	Medium	(7.982, 12.608, 15.692, 20.318)
	High	(15.692, 20.318, 23.401, 23.401)
Portuguese	Low	(1.2, 1.2, 5.792, 12.679)
	Medium	(5.792, 12.679, 17.271, 24.158)
	High	(17.271, 24.158, 28.75, 28.75)

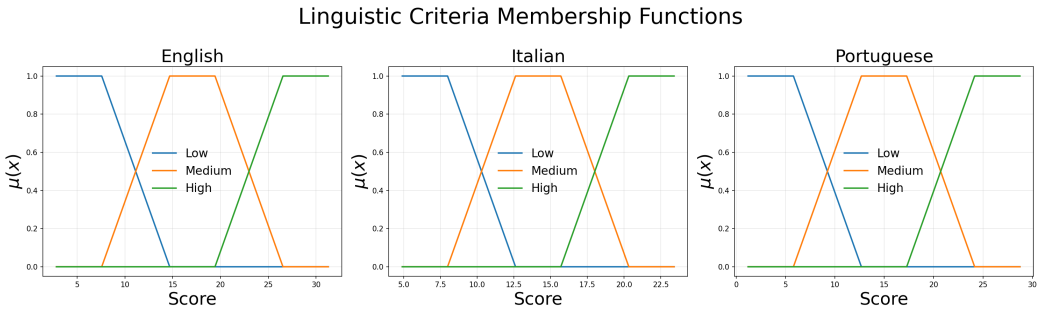

Figure 6. Trapezoidal membership functions used for the English, Italian, and Portuguese score variables.

Table 7. Resource-side criteria membership function parameters. Each row corresponds to a trapezoidal fuzzy number and its parameters.

Linguistic Variable	Fuzzy Set	Parameters
VRAM	Low	(8, 8, 16, 20)
	Medium	(16, 20, 24, 28)
	High	(24, 28, 32, 32)
Open-Sourceness	Low	(0, 0, 3, 5)
	Medium	(3, 5, 7, 9)
	High	(7, 9, 10, 10)

In particular, `VRAM` is modelled to reflect the idea of what is considered acceptable in relation to the price, based on the GPU's capacity. On the other hand, `Open-Sourceness` is described through its qualitative score range, which is sufficient for a smooth separation between low, intermediate, and high openness.

LLM openness is best understood as a multidimensional property rather than a binary distinction between "open source" and "closed" models. Relevant dimensions include the availability of model weights, which permits local execution but does not necessarily imply reproducibility; the release of source code, training pipelines, and, in stronger cases, training data under an OSI-compatible licence; disclosure or publication of pretraining and fine-tuning corpora, enabling auditability and replication; documentation of the model architecture and design choices through papers or technical reports; and the legal permissions granted by the model licence, including rights to commercial use, modification, and redistribution. Additional axes include open access through public APIs, which provides usability without revealing weights or internal implementation; open evaluation, in which

benchmark results, safety assessments, and testing protocols are published; and open fine-tuning or customisation, which allows users to adapt the model through methods such as LoRA or full parameter fine-tuning. Under the "gradient of openness" perspective, these dimensions are independent: a system may be open along some axes while remaining closed along others. For example, some models release weights but restrict licensing and withhold training data, whereas more fully open systems release weights, code, data documentation, and evaluation artefacts; proprietary API-only systems, by contrast, generally provide access without substantive openness of weights, data, or training methodology. In assigning this score, the stability over time and the clearness of the licence assigned to each model have been taken into account as concluding factors of the trustfulness of the model.

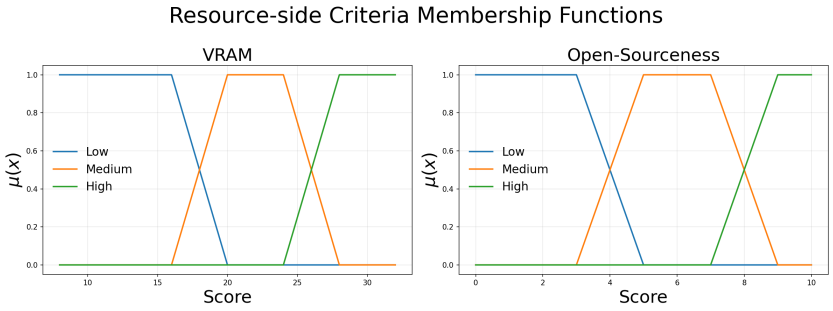

Figure 7. Membership functions used for the resource-side criteria `VRAM` and `Open-Sourceness`.

Finally, the output variable `Weight` is partitioned into five fuzzy numbers `Very Low`, `Low`, `Medium`, `High`, and `Very High` (see Table 8 and Figure 8). This latter partition is denser than the input ones, which is useful at the aggregation stage because it provides a more expressive scale for translating rule consequents into final scores.

Table 8. Output membership functions for the weight variable. Each row corresponds to a trapezoidal fuzzy number and its parameters.

Linguistic Variable	Fuzzy Set	Parameters
Weight	Very Low	(0, 0, 1, 2.25)
	Low	(1, 2.25, 3.25, 4.5)
	Medium	(3.25, 4.5, 5.5, 6.75)
	High	(5.5, 6.75, 7.75, 9)
	Very High	(7.75, 9, 10, 10)

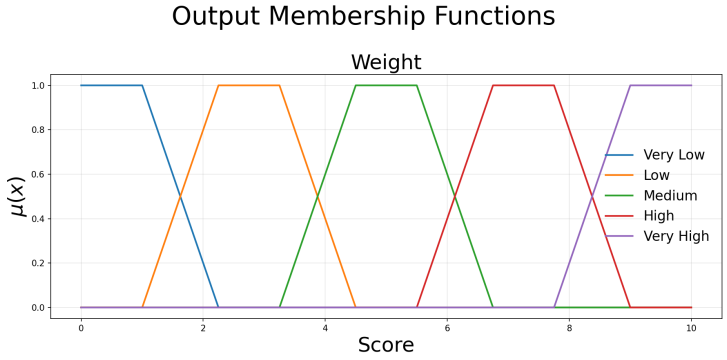

Figure 8. Membership functions used for the output variable `Weight`.

For the inference part, we consider the probabilistic t-norm $\otimes_P$ and t-conorm $\oplus_P$ for the interpretation of the logical operations `AND`, and `OR`, respectively.

The rule bases used in Sections 5.3 and 5.4 are authored manually to encode two contrasting, illustrative weighting policies; they are not learned from data. This is consistent

with the role of the FRBS in our framework: the rule base is the explicit, auditable statement of the decision-maker's preference structure (cf. [27]). Rule-based learning techniques such as Wang–Mendel [61] or genetic fuzzy systems [26] can be applied when historical decisions are available; we leave this hybrid setting for future work (Section 6).

5.3. Example 1: Language-Focused Weighting Policy

As a first working example for our inference system, we consider the following rule base:

R_1	IF `Italian` is `Low` AND `Portuguese` is `Low` AND `English` is `Low` THEN `Weight` is `Very Low`
R_2	IF `Italian` is `Low` OR `Portuguese` is `Low` OR `English` is `Low` THEN `Weight` is `Low`
R_3	IF `Italian` is `Medium` OR `Portuguese` is `Medium` OR `English` is `Medium` THEN `Weight` is `Medium`
R_4	IF `Italian` is `High` OR `Portuguese` is `High` OR `English` is `High` THEN `Weight` is `High`
R_5	IF `Italian` is `High` AND `Portuguese` is `High` AND `English` is `High` THEN `Weight` is `Very High`

This first example uses only linguistic inputs to assign a final weight. The idea is simple: strong competence in `Italian` and `Portuguese` is valued because both languages reflect a solid background in Latin. `English` is also useful, but, in this case, it plays a secondary role and can only partially compensate for weaker performance in the other two languages.

Each rule reflects a specific preference. When both `Italian` and `Portuguese` are `High`, the candidate receives a `Very High` weight because strong performance in both Latin languages is especially desirable. If `Portuguese` remains high while `Italian` is only `Medium`, the result is still `High`, showing that `Portuguese` carries slightly greater importance. When both Latin-language scores are `Low`, a high `English` score can still recover part of the evaluation, leading to a `Medium` weight. By contrast, a strong imbalance between `Italian` and `Portuguese` leads to a `Very Low` weight, because the model penalises inconsistent competence across the two key languages.

In Table 9, we show the final ranking for this first example, where the normalised score is the L_1 normalisation of the raw score. In this particular case study, `nemotron-cascade-2:30b` stands out clearly from the other candidates, with the highest raw score and a much higher normalised value. `mistral-small3.2:24b` follows at a moderate distance, while `gemma4:e4b` and `apertus:8b` receive substantially lower scores. This distribution shows that, under the language-focused rules of the first scenario, the fuzzy system is able to identify a strong preference for models that better match the desired Latin language profile.

Table 9. Final ranking for first example. The model `nemotron-cascade-2:30b` is the best under this scenario. The normalised scores are given by the L_1 normalisation of the row scores.

Rank	Model	Raw Score	Normalised Score
1	`nemotron-cascade-2:30b`	2.297	0.548
2	`mistral-small3.2:24b`	1.261	0.301
3	`gemma4:e4b`	0.624	0.149
4	`apertus:8b`	0.012	0.003

5.4. Example 2: Combined Linguistic and Resource-Aware Policy

In this second working example, we consider the following rule base:

R_1	IF Italian is Medium AND Portuguese is Medium AND English is Medium AND VRAM is Low THEN Weight is High
R_2	IF Italian is Medium AND Portuguese is Medium AND English is Medium AND Open-Sourceness is High THEN Weight is Very High
R_3	IF Italian is High AND Portuguese is Medium AND English is Medium THEN Weight is High
R_4	IF Italian is Low AND Portuguese is Medium AND English is Medium AND VRAM is High AND Open-Sourceness is Low THEN Weight is Very Low
R_5	IF English is High AND Open-Sourceness is Medium AND VRAM is Medium THEN Weight is Low

This second example extends the evaluation by including resource-related criteria in addition to language skills. Alongside Italian, Portuguese, and English, the model now also considers available VRAM and the degree of commitment to open-source software. This allows the final weight to reflect not only linguistic suitability, but also practical and ethical constraints.

Each rule highlights a different balance between these factors. When all three languages are at a Medium level, and VRAM is Low, the weight is still High because limited hardware resources make efficiency especially valuable. If the language profile remains balanced and Open-Sourceness is High, the weight becomes Very High, reflecting a strong preference for principled technical choices. A strong score in Italian, combined with Medium values in Portuguese and English, also produces a High weight because Italian is given special importance in this example. On the other hand, if Italian is Low, VRAM is High, and weak open-source alignment leads to a Very Low weight, since the candidate performs poorly on several key criteria. Finally, strong English alone is not enough to compensate for only Medium technical and ethical values, so that case receives a Low weight.

Table 10 summarises the final ranking produced in the second scenario, where the normalised score is again the L_1 normalisation of the row score. The results suggest that gemma4:e4b is the most suitable model under these preferences, with a clear lead in both raw and normalised score. The remaining models form a closer group, with apertus:8b, mistral-small3.2:24b, and nemotron-cascade-2:30b receiving lower but still comparable evaluations. Overall, the table shows how the fuzzy rule base can separate candidates according to the combined influence of language quality, resource efficiency, and open-source alignment.

Table 10. Final ranking for the second case. The model gemma4:e4b is the best under this scenario. The normalised scores are given by the L_1 normalisation of the row scores.

Rank	Model	Raw Score	Normalised Score
1	gemma4:e4b	2.592	0.360
2	apertus:8b	1.631	0.227
3	mistral-small3.2:24b	1.598	0.222
4	nemotron-cascade-2:30b	1.373	0.191

5.5. Sensitivity Analysis

In this section, we perform the sensitivity analysis of our method across the two examples provided in Sections 5.3 and 5.4.

5.5.1. Sensitivity Analysis of Example 1 in Section 5.3

The main idea to prove robustness is to delete each rule one time and to observe the effect. What should happen is that the output of this new, reduced set of rules should again

be reasonably aligned to the new user intent. This is slightly different from stating that by deleting any rule, the output should stay the same.

The reason why this should not be the case is the concept of the *Rules as material translation of the user's intent into the algorithm*. The core concept is that when a new rules set is created by deleting a rule, it represents a new (and potentially completely different) user's intent. By making a simple example, if a rule states that a particular coefficient should be a priority, this coefficient in an ideal system should not be a priority anymore if such a rule has been deleted (unless a redundant rule still exists).

In Example 1, since most of the rules are not significantly different, a user could expect that deleting any one of them would not produce significant changes in the overall ranking. This can be seen by comparing the full-system ranking in Tables 11 and 12 with the leave-one-rule-out rankings in Table 13. In these tables, there is a consistent ranking despite rule deletions, suggesting that the first intuition was right.

Table 11. Raw scores after ablating each rule for Example 1.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
Rule 1 ablated	1.662	0.929	0.579	0.015
Rule 2 ablated	1.256	0.772	0.239	0.016
Rule 3 ablated	3.022	1.602	0.820	0.016
Rule 4 ablated	2.740	1.507	0.730	0.015
Rule 5 ablated	2.778	1.486	0.735	0.000

Table 12. Normalised scores after ablating each rule for Example 1.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
Rule 1 ablated	0.522	0.292	0.182	0.005
Rule 2 ablated	0.550	0.338	0.105	0.007
Rule 3 ablated	0.553	0.293	0.150	0.003
Rule 4 ablated	0.549	0.302	0.146	0.003
Rule 5 ablated	0.556	0.297	0.147	0.000

Table 13. Ranks after ablating each rule for Example 1.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
Rule 1 ablated	1	2	3	4
Rule 2 ablated	1	2	3	4
Rule 3 ablated	1	2	3	4
Rule 4 ablated	1	2	3	4
Rule 5 ablated	1	2	3	4

5.5.2. Sensitivity Analysis of Example 2 in Section 5.4

The second example is more complex, since each rule targets a different combination of scores. From the user's perspective, it is unlikely that removing a rule would leave the ranking unchanged. The discussion below therefore considers each rule ablation separately.

If the first rule is removed, the deleted rule concerns performance and VRAM usage. A fallback rule may still apply in this case. However, that rule is highly complex and, as noted in the "Limitations" section, the present method appears to exhibit limited sensitivity to rules with negative consequents. As one would expect, removing the first rule shifts the emphasis towards raw performance, thereby favouring the most performant models.

If the second rule is removed, `Open-Sourceness` ceases to be prioritised. In this setting, the method continues to favour compact and high-performing models, such as `gemma4:e4b`, while also extending an advantage to other larger and more performant models.

If the third rule is removed, no observable change in the ranking is produced, since the rule is not sufficiently specific to alter the outcome.

Similarly, removing the fourth rule does not lead to substantial changes in the overall ranking, owing both to the limited sensitivity to negative rules and to the high specificity of that rule.

Lastly, removing the fifth rule does not significantly affect the overall ranking, again because the rule is highly specific and the method remains weakly sensitive to negative consequents.

We refer the reader to Tables 14–16 for a tabular presentation of the raw scores, the normalised scores, and the final ranking obtained from the rule-ablation analysis for Example 2.

Table 14. Raw scores after ablating each rule for Example 2.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
Rule 1 ablated	1.772	2.062	1.239	1.071
Rule 2 ablated	1.327	1.909	2.651	1.076
Rule 3 ablated	0.653	0.373	2.584	1.975
Rule 4 ablated	1.640	1.909	3.094	1.947
Rule 5 ablated	1.456	1.690	3.344	2.104

Table 15. Normalised scores after ablating each rule for Example 2.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
Rule 1 ablated	0.288	0.336	0.202	0.174
Rule 2 ablated	0.191	0.274	0.381	0.155
Rule 3 ablated	0.117	0.067	0.463	0.354
Rule 4 ablated	0.191	0.222	0.360	0.227
Rule 5 ablated	0.169	0.197	0.389	0.245

Table 16. Ranks after ablating each rule for Example 2.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
Rule 1 ablated	2	1	3	4
Rule 2 ablated	3	2	1	4
Rule 3 ablated	3	4	1	2
Rule 4 ablated	4	3	1	2
Rule 5 ablated	4	3	1	2

5.5.3. Sensitivity Analysis with Respect to t-norms

Tables 17–19 reports the results of Example 1 obtained under different t-norms. More precisely, since the reference values in this example are well separated, changing the t-norm does not produce any variation in the final ranking.

Likewise, Tables 20–22 presents the results of Example 2 for different t-norms. In this case, the choice of t-norm has a more pronounced effect, particularly under Łukasiewicz semantics, since the values are closer to one another and rank reversals are therefore more likely to occur.

Table 17. Raw scores obtained using product, Göedel and Łukasiewicz t-norms for Example 1.

t-norm	nemotron-cascade-2:30b	mistral-small3.2:24b	gemma4:e4b	apertus:8b
$\otimes_P$	2.297	1.261	0.624	0.012
$\otimes_G$	2.579	1.707	0.929	0.017
$\otimes_L$	2.183	0.990	0.442	0.000

Table 18. Normalised scores obtained using product, Göedel and Łukasiewicz t-norms for Example 1.

t-norm	nemotron-cascade-2:30b	mistral-small3.2:24b	gemma4:e4b	apertus:8b
$\otimes_P$	0.548	0.301	0.149	0.003
$\otimes_G$	0.493	0.326	0.178	0.003
$\otimes_L$	0.604	0.274	0.122	0.000

Table 19. Ranks obtained using product, Göedel and Łukasiewicz t-norms for Example 1.

t-norm	nemotron-cascade-2:30b	mistral-small3.2:24b	gemma4:e4b	apertus:8b
$\otimes_P$	1	2	3	4
$\otimes_G$	1	2	3	4
$\otimes_L$	1	2	3	4

Table 20. Raw scores obtained using the product, Gödel, and Łukasiewicz t-norms for Example 2.

	nemotron-cascade-2:30b	mistral-small3.2:24b	gemma4:e4b	apertus:8b
$\otimes_P$	1.373	1.598	2.592	1.631
$\otimes_G$	2.037	1.598	2.592	2.411
$\otimes_L$	0.832	1.598	2.592	1.185

Table 21. Normalised scores obtained using the product, Gödel, and Łukasiewicz t-norms for Example 2.

	nemotron-cascade-2:30b	mistral-small3.2:24b	gemma4:e4b	apertus:8b
$\otimes_P$	0.191	0.222	0.360	0.227
$\otimes_G$	0.236	0.185	0.300	0.279
$\otimes_L$	0.134	0.258	0.418	0.191

Table 22. Ranks obtained using the product, Gödel, and Łukasiewicz t-norms for Example 2.

	nemotron-cascade-2:30b	mistral-small3.2:24b	gemma4:e4b	apertus:8b
$\otimes_P$	4	3	1	2
$\otimes_G$	3	4	1	2
$\otimes_L$	4	2	1	3

5.5.4. Rank Comparison Using Kendall τ

A final robustness check was carried out by perturbing the input scores and assessing the stability of the resulting model rankings. More precisely, the mean values of the trapezoidal numbers associated with the input scores were perturbed by a percentage amount, reported on the x-axis of Figures 9 and 10. For each perturbation level, 100 independent runs were performed. Each perturbed ranking was then compared with the original ranking by means of Kendall's τ, computed on the ordering of the models. For

each set of 100 runs, the mean and standard deviation of Kendall's τ [68] were computed and represented on the y-axis.

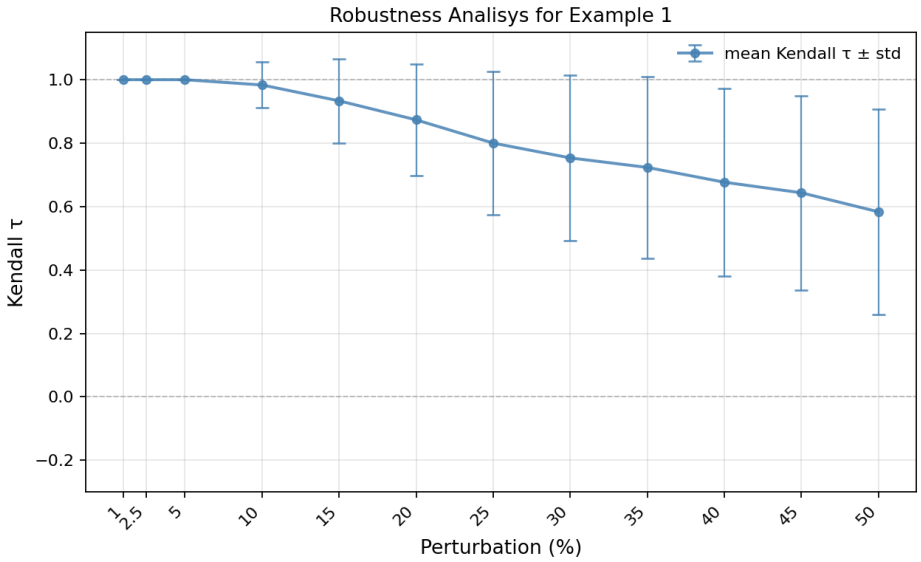

Figure 9. Robustness analysis for Example 1. The curve reports the mean Kendall's τ over 100 runs for each perturbation level, with error bars indicating one standard deviation.

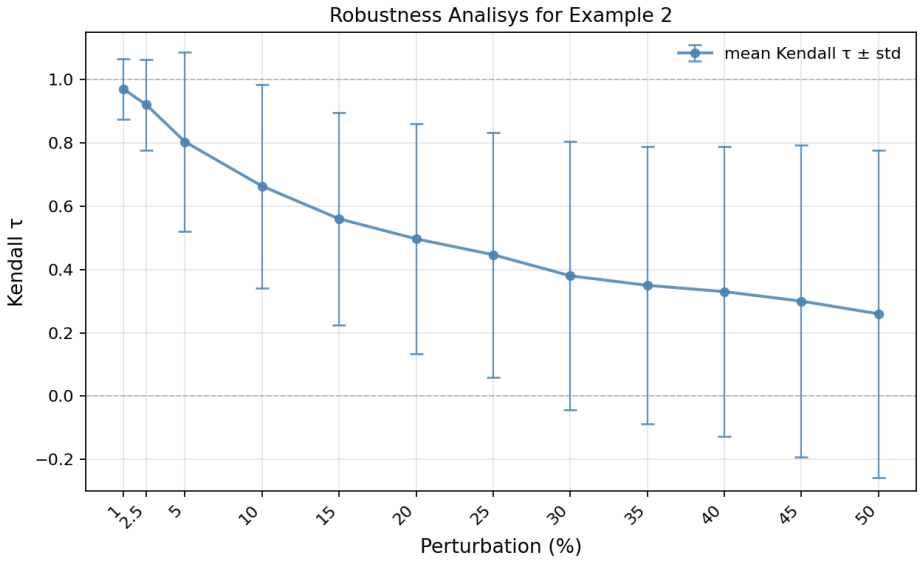

Figure 10. Robustness analysis for Example 2. The curve reports the mean Kendall's τ over 100 runs for each perturbation level, with error bars indicating one standard deviation.

This analysis provides a quantitative assessment of the robustness of the method in the two examples. In Example 1, shown in Figure 9, the ranking remains highly stable under small perturbations and declines only gradually as the perturbation level increases. Even for larger perturbations, the average Kendall's τ remains positive, indicating that the ranking is only partially affected.

In Example 2, shown in Figure 10, the ranking exhibits greater sensitivity to perturbations. The mean Kendall's τ decreases more rapidly than in Example 1, while the standard deviation increases as the perturbation level grows. This suggests that Example 2 is less robust under score perturbation. Such behaviour is consistent with the previous analyses:

the rules in Example 2 encode more differentiated criteria, and the model scores are closer to one another, thereby making rank reversals more likely.

Overall, the perturbation analysis confirms that Kendall's τ is a useful indicator of ranking stability. At the same time, it should be interpreted in conjunction with the semantic role of the rules. In particular, a lower value of τ does not necessarily signal an error, but rather indicates that the induced ranking is sensitive to perturbations in the underlying scores or in the contribution of the rules.

5.6. Comparison with Existing Methods

The following tables compare the proposed fuzzy rule-based method with three alternative decision-making frameworks, namely Markov Logic Networks (MLNs), ProbLog, and TOPSIS. All methods are evaluated on the same two examples and on the same set of models, although they differ substantially in the way they encode both the input information and the rule structure.

The reference approach adopted throughout the previous sections is a Larsen-type fuzzy rule-based system. In this framework, the raw model scores are represented by trapezoidal fuzzy numbers, and each linguistic condition is evaluated through its corresponding membership degree. Each rule is activated with a graded firing strength, computed by applying the product *t*-norm to its antecedents, while the final score is obtained by aggregating the consequents and subsequently defuzzifying the result. Consequently, the reference method preserves both the linguistic form of the rule base and the graded character of the input data.

The comparison with Markov Logic Networks (MLNs) [69] is obtained by translating the same IF–THEN rules into weighted logical clauses. Under this interpretation, each rule contributes to a joint log-linear probability distribution over possible worlds. The input membership values are transformed into unary log-odds potentials, and each rule is assigned a fixed reliability weight. This makes MLNs a natural probabilistic-logic baseline, but it also changes the semantics of the reasoning process: rules no longer fire independently with a fuzzy degree, but instead interact globally through a probabilistic model. The resulting MLN scores tend to be more compressed after normalisation, especially in Example 1, where all models receive relatively similar normalised values.

ProbLog [70] provides a second comparison grounded in probabilistic logic. In this setting, the fuzzy membership values are used directly as probabilities assigned to probabilistic facts, while the IF–THEN rules are encoded as deterministic Prolog clauses. The probability of each output label is then computed by exact probabilistic inference over possible worlds, and the final score is derived from the expected value of the output labels. In contrast to MLNs, ProbLog does not require the introduction of an additional rule-weight parameter. Overall, ProbLog yields results that are very close to those of the reference fuzzy method. In Example 1, the similarity concerns both the final ordering and the distribution of the normalised scores. In Example 2, ProbLog likewise preserves the same ordering as the reference system, suggesting that it is the closest alternative among the three baselines considered here.

TOPSIS [15] follows a substantially different rationale. It is not a rule-based approach, but rather a multi-criteria decision-making method that ranks alternatives according to their distance from an ideal and an anti-ideal solution. Unlike the proposed fuzzy method, TOPSIS operates directly on the raw mean scores and does not employ membership functions, rule antecedents, consequents, or defuzzification. The orientation of each criterion must therefore be specified explicitly. In Example 1, the language scores are treated as benefit criteria, whereas in Example 2, language performance and `Open-Sourceness` are treated as benefit criteria, and `VRAM` is treated as a cost criterion. TOPSIS is able to reproduce

the ordering of Example 1, but diverges in Example 2, since it captures only criterion-wise preference directions and cannot encode the conditional rule structure specified by the user.

Taken together, these comparisons indicate that the proposed fuzzy method occupies an intermediate position between symbolic rule-based reasoning and purely numerical ranking procedures. ProbLog is the closest alternative, because it preserves the logical structure of the rule base while interpreting membership values probabilistically. MLNs also retain the rule structure, but embed it within a global weighted probabilistic semantics. By contrast, TOPSIS offers a simpler and more transparent multi-criteria baseline, yet it cannot directly express conditional IF–THEN preferences.

The complete numerical results for both examples are reported in Tables 23–28.

Table 23. Raw scores obtained using MLN and ProbLog for Example 1.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
MLN	13.801	12.101	11.565	11.141
ProbLog	11.098	6.105	2.877	0.070

Table 24. Normalised scores obtained using MLN, ProbLog, and TOPSIS for Example 1.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
MLN	0.284	0.249	0.238	0.229
ProbLog	0.551	0.303	0.143	0.003
TOPSIS	0.974	0.710	0.506	0.000

Table 25. Ranks obtained using MLN, ProbLog, and TOPSIS for Example 1.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
MLN	1	2	3	4
ProbLog	1	2	3	4
TOPSIS	1	2	3	4

Table 26. Raw scores obtained using MLN and ProbLog for Example 2.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
MLN	11.643	12.275	13.562	11.502
ProbLog	6.551	7.104	9.537	8.087

Table 27. Normalised scores obtained using MLN, ProbLog, and TOPSIS for Example 2.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
MLN	0.238	0.251	0.277	0.235
ProbLog	0.209	0.227	0.305	0.259
TOPSIS	0.614	0.503	0.574	0.449

Table 28. Ranks obtained using MLN, ProbLog, and TOPSIS for Example 2.

	nemotron-cascade-2:30b	**mistral-small3.2:24b**	**gemma4:e4b**	**apertus:8b**
MLN	3	2	1	4
ProbLog	4	3	1	2
TOPSIS	1	3	2	4

5.7. End-to-End MCGDM Aggregation Example

In this section, apply the standard weighted MCGDM formulation in order to illustrate how the expert weights w_k produced by the FRBS can be used in a concrete group decision-making problem. The problem instance consists of a set of $m = 3$ *alternatives*, $A = \{A_1, A_2, A_3\}$, a set of $n = 3$ *criteria*, $C = \{C_1, C_2, C_3\}$, and a group of $K = 4$ *experts*, $D = \{D_1, D_2, D_3, D_4\}$. The criteria are $C_1 =$ Usefulness, $C_2 =$ Factual Correctness, and $C_3 =$ Text Quality, each assessed on an integer scale from 0 to 10. Criterion weights are assumed to be uniform, namely $v_j = \frac{1}{3}$ for all j.

The four experts are the LLMs considered in Section 5.4, with weight vector $\mathbf{w} = (w_1, w_2, w_3, w_4)$ satisfying

$$\sum_{k=1}^{K} w_k = 1, \qquad w_k \geq 0, \tag{28}$$

and taken from Table 10. More precisely,

$$\mathbf{w} = (0.003,\ 0.149,\ 0.301,\ 0.548)$$

for `apertus:8b`, `gemma4:e4b`, `mistral-small3.2:24b`, and `nemotron-cascade-2:30b`, respectively.

The three alternatives correspond to actual generated answers to the question *"What is the current state of the European energy transition?"* Specifically, A_1 is a factually correct answer in English, A_2 is a factually incorrect answer in Italian, and A_3 is the English translation of A_2. Each expert D_k independently evaluates every alternative A_i with respect to every criterion C_j, thereby producing an individual score $x_{ij}^{(k)}$. The resulting per-expert evaluations, obtained by directly querying the models, are reported in Table 29.

The aggregation is carried out in two stages. In Step 1, the group evaluation of alternative A_i with respect to criterion C_j is computed as the expert-weighted average of the individual scores:

$$\bar{x}_{ij} = \sum_{k=1}^{K} w_k \, x_{ij}^{(k)}, \tag{29}$$

thus yielding the aggregated decision matrix $\bar{X} = [\bar{x}_{ij}]$.

In Step 2, the aggregated matrix $\bar{X}$ is reduced to a scalar score for each alternative by means of the criterion weights:

$$S_i = \sum_{j=1}^{n} v_j \, \bar{x}_{ij} = \sum_{j=1}^{n} \sum_{k=1}^{K} v_j \, w_k \, x_{ij}^{(k)}, \tag{30}$$

and the preferred alternative is then selected according to

$$A^* = \arg\max_i \ S_i. \tag{31}$$

Table 30 reports the aggregated matrix $\bar{X}$ together with the final scores S_i obtained from Equations (29) and (30).

The method selects $A^* = A_1$ with $S_1 = 8.88$, correctly identifying the factually accurate answer. The dominant discriminator is C_2 (Factual Correctness): both A_2 and A_3 receive markedly lower scores on this criterion from the two highest-weighted experts (`nemotron-cascade-2:30b` and `mistral-small3.2:24b`), which propagates into a substantial gap in $\bar{x}_{i2}$ (8.61 vs. 5.32 and 4.32) and ultimately in S_i. The ranking $A_1 \succ A_2 \succ A_3$ holds under any positive assignment of criterion weights, since A_1 dominates on every $\bar{x}_{ij}$.

Table 29. Per-expert scores $x_{ij}^{(k)}$ (integer, scale 0–10) and per-expert WSM $x_i^{(k)} = \frac{1}{3}\sum_j x_{ij}^{(k)}$ obtained by querying each LLM on the three answer alternatives.

Expert (w_k)		C_1	C_2	C_3
`nemotron-cascade-2:30b` (0.548)	A_1	9	8	9
	A_2	8	5	7
	A_3	7	4	7
`mistral-small3.2:24b` (0.301)	A_1	9	9	9
	A_2	6	4	7
	A_3	7	4	6
`gemma4:e4b` (0.149)	A_1	9	10	9
	A_2	9	9	10
	A_3	9	6	9
`apertus:8b` (0.003)	A_1	10	9	10
	A_2	9	10	8
	A_3	9	10	9

Table 30. Aggregated decision matrix $\bar{x}_{ij} = \sum_k w_k x_{ij}^{(k)}$ and final scores $S_i = \frac{1}{3}\sum_j \bar{x}_{ij}$. The best alternative A^* is the one maximising S_i.

Alternative	$\bar{x}_{i1}$	$\bar{x}_{i2}$	$\bar{x}_{i3}$	S_i
A_1	9.01	8.61	9.01	**8.88**
A_2	7.56	5.32	7.46	6.78
A_3	7.31	4.32	7.01	6.21

6. Conclusions and Future Works

This work has examined the use of a Fuzzy Rule-Based System for assigning weights to experts within a Multi-Criteria Group Decision-Making framework, with particular attention to settings in which artificial experts contribute alongside, or in place of, human decision-makers. Building on the preliminary methodological proposal introduced in [10], the present contribution has focused on an application-oriented validation of the approach, showing how an interpretable fuzzy weighting mechanism can be adapted to a concrete decision domain and how its behaviour changes under different modelling choices.

The results obtained from the LLM case study confirm that the proposed framework is sufficiently flexible to encode heterogeneous weighting policies through transparent IF–THEN rules. In particular, the experiments have shown that the final weights are not determined by performance alone, but may be shaped in a controlled and auditable manner by combining multilingual quality, computational requirements, and openness-related characteristics. This is precisely one of the main advantages of the FRBS perspective: rather than imposing a single optimisation principle, it allows the weighting scheme to reflect the priorities of the decision context. In this sense, the method appears well-suited to real-world scenarios, where decision quality is rarely reducible to one isolated dimension.

The empirical examples also highlight an important substantive point. Small changes in the rule base, or in the logical structure of the antecedents, may produce substantial changes in the final ranking. This should not be viewed as a weakness of the method, but rather as evidence of its sensitivity to explicit modelling assumptions. Hence, the framework makes the value judgements underlying the weighting process visible, instead of leaving them implicit in a black-box optimisation step. From the perspective of MCGDM, this is a significant methodological advantage, especially when artificial experts are involved, and accountability is required.

At the same time, the study has shown that some modelling components are particularly influential. Among them, the design of the rule base is the most decisive, since it directly encodes the weighting policy adopted in the application. The choice of membership functions and the partition of the linguistic variables also affect the outcome, especially in borderline cases, where an expert lies near the overlap of adjacent fuzzy sets. Likewise, the selection of logical operators and the adopted defuzzification procedure may alter the relative magnitude of the final scores. For this reason, the proposed framework should not be interpreted as a fully automatic mechanism, but as a structured and interpretable tool whose reliability depends on a careful alignment between fuzzy modelling choices and domain-specific objectives.

From a practical standpoint, the proposed approach offers a viable way of integrating artificial experts into decision-support pipelines without treating them as interchangeable black boxes. By assigning weights through linguistic variables and interpretable rules, the framework supports a more nuanced form of aggregation in which the role of each expert can be justified explicitly. This is particularly relevant in contemporary decision environments, where algorithmic agents, recommender systems, predictive models, and large language models increasingly act as decision contributors rather than as mere auxiliary tools.

The present study nevertheless has some limitations. The empirical validation has been conducted on a single case study involving four LLMs evaluated on three multilingual criteria and two resource-side criteria. Although this design suffices to expose the operational behaviour of the framework under different rule bases (Sections 5.3 and 5.4) and to provide a controlled setting for sensitivity analysis (Section 5.5) and a comparison with other existing methods (Section 5.6), the conclusions should not be overgeneralised; a broader benchmark over multiple decision domains and larger expert pools is left for future work.

Several lines of future research follow naturally from these observations. A first direction concerns the extension of the experimental analysis to additional application domains, including more traditional MCGDM settings and larger groups of heterogeneous experts. A second direction is the development of a more systematic sensitivity analysis covering membership-function design, logical connectives, inference schemes, and defuzzification operators. A third possible development is the introduction of semi-automatic or data-assisted procedures for constructing the rule base, so as to reduce the burden of manual design while preserving interpretability. Finally, it would be worthwhile to investigate hybrid settings in which FRBS-based weighting is combined with empirical validation, calibration procedures, or benchmark-based evaluation protocols for artificial experts.

In conclusion, the main contribution of this paper lies not in proposing yet another abstract weighting formula, but in showing that expert weighting can be formulated as an interpretable fuzzy inference problem and meaningfully instantiated in an application-oriented setting. The study, therefore, strengthens the case for FRBS-based weighting as a transparent and operationally credible approach for MCGDM, especially in scenarios where human and artificial expertise must be combined under multiple, and potentially competing, criteria.

Author Contributions: L.C.: Methodology, Formal analysis, Visualization, Writing—Original Draft, Writing—Review and Editing. G.F.: Conceptualization, Methodology, Formal analysis, Visualization, Writing—Original Draft, Writing—Review and Editing. G.G.: Methodology, Formal analysis, Visualization, Writing—Original Draft, Writing—Review and Editing, Software, Validation, Investigation. G.L.R.: Methodology, Formal analysis, Visualization, Writing—Original Draft, Writing—Review and Editing. M.E.T.: Supervision, Project administration, Funding acquisition, Writing—Review and Editing. All authors have read and agreed to the published version of the manuscript.

Funding: Lydia Castronovo acknowledges support by the **Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni** (**GNAMPA**) of the **Istituto Nazionale di Alta Matematica** (**INdAM**) "Francesco Severi". Giuseppe Filippone, Giuseppe Giacopelli, Gianmarco La Rosa, and Marco Elio Tabacchi acknowledge financial support from the **Sustainability Decision Framework** (**SDF**) Research Project—CUP **B79J23000540005**—Grant Assignment Decree No. **5486** adopted on **2023-08-04**. Gianmarco La Rosa also acknowledges support from the **Gruppo Nazionale per le Strutture Algebriche, Geometriche e le loro Applicazioni** (**GNSAGA**) of the **Istituto Nazionale di Alta Matematica** (**INdAM**) "Francesco Severi". Lydia Castronovo and Marco Elio Tabacchi also acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 1409 published on 14.9.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union—NextGenerationEU—Project Title **Quantum Models for Logic, Computation and Natural Processes** (**QM4NP**)—CUP **B53D23030160001**—Grant Assignment Decree No. **1371** adopted on **2023-09-01** by the Italian Ministry of University and Research (MUR).

Data Availability Statement: The data and Python v3.10 scripts supporting this manuscript are available in the GitHub repository at https://github.com/SDF-Unipa/logic_operations_in_a_frbs, (accessed on 28 April 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

FRBS	Fuzzy Rule-Based System
DM	Decision Maker
MCDM	Multi-Criteria Decision-Making
MCGDM	Multi-Criteria Group Decision-Making
WSM	Weighted Sum Method
XAI	Explainable Artificial Intelligence
FCM	Fuzzy C-Means
FST	Fuzzy Set Theory
CoG	Centre of Gravity
MOM	Mean of Maximum
LLM	Large Language Model
AI	Artificial Intelligence
GMP	Generalised Modus Ponens
CRI	Compositional Rule of Inference
UoD	Universe of Discourse
TFN	Triangular Fuzzy Number
TpFN	Trapezoidal Fuzzy Number
SDF	Sustainability Decision Framework
GNAMPA	Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni
INdAM	Istituto Nazionale di Alta Matematica
GNSAGA	Gruppo Nazionale per le Strutture Algebriche, Geometriche e le loro Applicazioni
NRRP	National Recovery and Resilience Plan
MUR	Ministry of University and Research
QM4NP	Quantum Models for Logic, Computation and Natural Processes

References

1. Boix-Cots, D.; Pardo-Bosch, F.; Pujadas, P. A systematic review on multi-criteria group decision-making methods based on weights: Analysis and classification scheme. *Inf. Fusion* **2023**, *96*, 16–36. https://doi.org/10.1016/j.inffus.2023.03.004.
2. Saaty, T.L. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*; McGraw-Hill International Book Co.: New York, NY, USA; London, UK, 1980.
3. Uzhga-Rebrov, O.; Kuļešova, G. A Review and Comparative Analysis of Methods for Determining Criteria Weights in MCDM Tasks. *Inf. Technol. Manag. Sci.* **2023**, *26*, 35–40. https://doi.org/10.7250/itms-2023-0005.

4. Du, Y.W.; Zhong, J.J. Dynamic multicriteria group decision-making method with automatic reliability and weight calculation. *Inf. Sci.* **2023**, *634*, 400–422. https://doi.org/10.1016/j.ins.2023.03.092.
5. Tapia Garcı´a, J.; del Moral, M.; Martínez, M.; Herrera-Viedma, E. A consensus model for group decision making problems with linguistic interval fuzzy preference relations. *Expert Syst. Appl.* **2012**, *39*, 10022–10030. https://doi.org/10.1016/j.eswa.2012.02.008.
6. Liao, H.; Xu, Z.; Zeng, X.J.; Xu, D.L. An enhanced consensus reaching process in group decision making with intuitionistic fuzzy preference relations. *Inf. Sci.* **2016**, *329*, 274–286. https://doi.org/10.1016/j.ins.2015.09.024.
7. Ayan, B.; Abacıoğlu, S.; Basilio, M.P. A Comprehensive Review of the Novel Weighting Methods for Multi-Criteria Decision-Making. *Information* **2023**, *14*, 285. https://doi.org/10.3390/info14050285.
8. Kostopoulos, G.; Davrazos, G.; Kotsiantis, S. Explainable Artificial Intelligence-Based Decision Support Systems: A Recent Review. *Electronics* **2024**, *13*, 2842. https://doi.org/10.3390/electronics13142842.
9. Soori, M.; Jough, F.K.G.; Dastres, R.; Arezoo, B. AI-based decision support systems in Industry 4.0, a review. *J. Econ. Technol.* **2026**, *4*, 206–225. https://doi.org/10.1016/j.ject.2024.08.005.
10. Castronovo, L.; Filippone, G.; La Rosa, G.; Tabacchi, M.E. Fuzzy Rule-Based Approach for Weighting Artificial Experts involved in a Multi-Criteria Group Decision-Making Problem. In *Proceedings of the 21st International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2026), Rome, Italy, 15–19 June 2026*; Communications in Computer and Information Science; Springer: Berlin/Heidelberg, Germany, 2016 Available online: https://link.springer.com/book/9783032289964 (accessed on 21 May 2026).
11. Tabacchi, M.E.; Termini, S., Experimental Modeling for a Natural Landing of Fuzzy Sets in New Domains. In *Enric Trillas: A Passion for Fuzzy Sets: A Collection of Recent Works on Fuzzy Logic*; Springer International Publishing: Cham, Switzerland, 2015; Volume 322, pp. 179–188. https://doi.org/10.1007/978-3-319-16235-5_13.
12. Tabacchi, M.E.; Portmann, E.; Seising, R.; Habenstein, A., Designing Cognitive Cities. In *Designing Cognitive Cities*; Springer International Publishing: Cham, Switzerland, 2019; Volume 176, pp. 3–27. https://doi.org/10.1007/978-3-030-00317-3_1.
13. Martin Larsen, P. Industrial applications of fuzzy logic control. *Int. J. Man-Mach. Stud.* **1980**, *12*, 3–10. https://doi.org/10.1016/S0020-7373(80)80050-2.
14. Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J.M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Díaz-Rodríguez, N.; Herrera, F. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Inf. Fusion* **2023**, *99*, 101805. https://doi.org/10.1016/j.inffus.2023.101805.
15. Hwang, C.; Yoon, K. *Multiple Attribute Decision Making. Methods and Applications A State-of-the-Art Survey*, 1st ed.; Lecture Notes in Economics and Mathematical Systems; Springer: Berlin/Heidelberg, Germany, 1981; pp. XI, 269. https://doi.org/10.1007/978-3-642-48318-9.
16. Opricovic, S. Programski paket VIKOR za visekriterijumsko kompromisno rangiranje. In Proceedings of the 17th International Symposium on Operational Research SYM-OP-IS, Kupari, Croatia, 9–12 October 1990.
17. Opricovic, S. Multi-criteria optimization of civil engineering systems. *Fac. Civ. Eng. Belgrade* **1998**, *2*, 5–21.
18. Opricovic, S.; Tzeng, G.H. Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *Eur. J. Oper. Res.* **2004**, *156*, 445–455. https://doi.org/10.1016/S0377-2217(03)00020-1.
19. Cooke, R.M. *Experts in Uncertainty: Opinion and Subjective Probability in Science*; Oxford University Press: New York, NY, USA, 1991.
20. Clemen, R.T.; Winkler, R.L. Combining Probability Distributions from Experts in Risk Analysis. *Risk Anal.* **1999**, *19*, 187–203. https://doi.org/10.1111/j.1539-6924.1999.tb00399.x.
21. Mamdani, E. Application of fuzzy algorithms for control of simple dynamic plant. *Proc. Inst. Electr. Eng.* **1974**, *121*, 1585–1588. https://doi.org/10.1049/piee.1974.0328.
22. Mamdani, E.; Assilian, S. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-Mach. Stud.* **1975**, *7*, 1–13. https://doi.org/10.1016/S0020-7373(75)80002-2.
23. Takagi, T.; Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.* **1985**, *SMC-15*, 116–132. https://doi.org/10.1109/TSMC.1985.6313399.
24. Dubois, D.; Prade, H. What are fuzzy rules and how to use them. *Fuzzy Sets Syst.* **1996**, *84*, 169–185. https://doi.org/10.1016/0165-0114(96)00066-8.
25. Cordón, O.; Herrera, F.; Hoffmann, F.; Magdalena, L., Fuzzy Rule-Based Systems. In *Genetic Fuzzy Systems*; WORLD SCIENTIFIC: Singapore, 2001; Volume 19, pp. 1–46. https://doi.org/10.1142/9789812810731_0001.
26. Cordón, O. A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems. *Int. J. Approx. Reason.* **2011**, *52*, 894–913. https://doi.org/10.1016/j.ijar.2011.03.004.
27. Mendel, J.M. *Explainable Uncertain Rule-Based Fuzzy Systems*, 3rd ed.; Springer: Cham, Switzerland, 2024; pp. XXIII, 580. https://doi.org/10.1007/978-3-031-35378-9.
28. Koksalmis, E.; Kabak, Ö. Deriving decision makers' weights in group decision making: An overview of objective methods. *Inf. Fusion* **2019**, *49*, 146–160.

29. Wang, J.; Wei, G.; Wei, C.; Wu, J. Maximizing deviation method for multiple attribute decision making under q-rung orthopair fuzzy environment. *Def. Technol.* **2020**, *16*, 1073–1087. https://doi.org/10.1016/j.dt.2019.11.007.
30. Peng, H.g.; Zhang, H.y.; Wang, J.q.; Li, L. An uncertain Z-number multicriteria group decision-making method with cloud models. *Inf. Sci.* **2019**, *501*, 136–154. https://doi.org/10.1016/j.ins.2019.05.090.
31. Mersha, M.; Lam, K.; Wood, J.; AlShami, A.K.; Kalita, J. Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. *Neurocomputing* **2024**, *599*, 128111. https://doi.org/10.1016/j.neucom.2024.128111.
32. Yang, W.; Wei, Y.; Wei, H.; Chen, Y.; Huang, G.; Li, X.; Li, R.; Yao, N.; Wang, X.; Gu, X.; et al. Survey on Explainable AI: From Approaches, Limitations and Applications Aspects. *Hum.-Centric Intell. Syst.* **2023**, *3*, 161–188. https://doi.org/10.1007/s44230-023-00038-y.
33. Jiang, J.; Liu, X.; Wang, Z.; Ding, W.; Zhang, S.; Xu, H. Large group decision-making with a rough integrated asymmetric cloud model under multi-granularity linguistic environment. *Inf. Sci.* **2024**, *678*, 120994. https://doi.org/10.1016/j.ins.2024.120994.
34. Tanino, T., Sensitivity Analysis in MCDM. In *Multicriteria Decision Making: Advances in MCDM Models, Algorithms, Theory, and Applications*; Springer: Boston, MA, USA, 1999; pp. 173–201. https://doi.org/10.1007/978-1-4615-5025-9_7.
35. Więckowski, J.; Sałabun, W. Sensitivity analysis approaches in multi-criteria decision analysis: A systematic review. *Appl. Soft Comput.* **2023**, *148*, 110915. https://doi.org/10.1016/j.asoc.2023.110915.
36. Zimmerman, H.J. *Fuzzy Set Theory and Its Applications*, 4th ed.; Springer: Dordrecht, The Netherlands, 1996; p. XXV, 514. https://doi.org/10.1007/978-94-010-0646-0
37. Zadeh, L. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X.
38. Zadeh, L.A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst.* **1997**, *90*, 111–127. https://doi.org/10.1016/S0165-0114(97)00077-8.
39. Dubois, D.; Prade, H. A review of fuzzy set aggregation connectives. *Inf. Sci.* **1985**, *36*, 85–121. https://doi.org/10.1016/0020-0255(85)90027-1.
40. Hájek, P. *Metamathematics of Fuzzy Logic*, 1st ed.; Trends in Logic; Springer: Dordrecht, The Netherland, 1998; pp. VIII, 299. https://doi.org/10.1007/978-94-011-5300-3.
41. Zadeh, L.A. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Trans. Syst. Man, Cybern.* **1973**, *SMC-3*, 28–44. https://doi.org/10.1109/TSMC.1973.5408575.
42. Buckley, J.J.; Eslami, E. *An Introduction to Fuzzy Logic and Fuzzy Sets*, 1st ed.; Advances in Intelligent and Soft Computing, Physica Heidelberg; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2002; pp. X, 285. https://doi.org/10.1007/978-3-7908-1799-7.
43. van Laarhoven, P.J.M.; Pedrycz, W. A fuzzy extension of Saaty's priority theory. *Fuzzy Sets Syst.* **1983**, *11*, 229–241. https://doi.org/10.1016/S0165-0114(83)80082-7.
44. Castronovo, L.; Filippone, G.; Galici, M.; La Rosa, G.; Tabacchi, M.E. Fuzzy MCGDM Approach for Ontology Fuzzification. *Electronics* **2025**, *14*, 3596. https://doi.org/10.3390/electronics14183596.
45. Triantaphyllou, E. *Multi-Criteria Decision Making Methods: A Comparative Study*; Springer: New York, NY, USA, 2000; Volume 44. https://doi.org/10.1007/978-1-4757-3157-6.
46. Simo, U.F.; Gwét, H. Fuzzy Triangular Aggregation Operators. *Int. J. Math. Math. Sci.* **2018**, *2018*, 9209524. https://doi.org/10.1155/2018/9209524.
47. Aliyeva, K.; Aliyeva, A.; Aliyev, R.; Özdeşer, M. Application of Fuzzy Simple Additive Weighting Method in Group Decision-Making for Capital Investment. *Axioms* **2023**, *12*, 797. https://doi.org/10.3390/axioms12080797.
48. Wang, Y.J. A fuzzy multi-criteria decision-making model based on simple additive weighting method and relative preference relation. *Appl. Soft Comput.* **2015**, *30*, 412–420. https://doi.org/10.1016/j.asoc.2015.02.002.
49. Nădăban, S.; Dzitac, S.; Dzitac, I. Fuzzy TOPSIS: A General View. *Procedia Comput. Sci.* **2016**, *91*, 823–831. https://doi.org/10.1016/j.procs.2016.07.088.
50. Le, H.T.; Chu, T.C. Ranking Alternatives Using a Fuzzy Preference Relation-Based Fuzzy VIKOR Method. *Axioms* **2023**, *12*, 1079. https://doi.org/10.3390/axioms12121079.
51. Turgut, Z.K.; Tolga, A.Ç., Fuzzy MCDM Methods in Sustainable and Renewable Energy Alternative Selection: Fuzzy VIKOR and Fuzzy TODIM. In *Energy Management—Collective and Computational Intelligence with Theory and Applications*; Springer International Publishing: Cham, Switzerland, 2018; pp. 277–314.
52. Zhang, C.; Ma, C.b.; Xu, J.d. A New Fuzzy MCDM Method Based on Trapezoidal Fuzzy AHP and Hierarchical Fuzzy Integral. In *Proceedings of the Fuzzy Systems and Knowledge Discovery*; Wang, L., Jin, Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 466–474. https://doi.org/10.1007/11540007_57.
53. Moslem, S.; Farooq, D.; Jamal, A.; Almarhabi, Y.; Almoshaogeh, M.; Butt, F.M.; Tufail, R.F. An Integrated Fuzzy Analytic Hierarchy Process (AHP) Model for Studying Significant Factors Associated with Frequent Lane Changing. *Entropy* **2022**, *24*, 367. https://doi.org/10.3390/e24030367.

54. Akram, M.; Luqman, A.; Kahraman, C. Hesitant Pythagorean fuzzy ELECTRE-II method for multi-criteria decision-making problems. *Appl. Soft Comput.* **2021**, *108*, 107479. https://doi.org/10.1016/j.asoc.2021.107479.
55. Chu, T.C.; Nghiem, T.B.H. Extension of Fuzzy ELECTRE I for Evaluating Demand Forecasting Methods in Sustainable Manufacturing. *Axioms* **2023**, *12*, 926. https://doi.org/10.3390/axioms12100926.
56. Senvar, O.; Tuzkaya, G.; Kahraman, C., Multi Criteria Supplier Selection Using Fuzzy PROMETHEE Method. In *Supply Chain Management Under Fuzziness: Recent Developments and Techniques*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 21–34. https://doi.org/10.1007/978-3-642-53939-8_2.
57. Papapostolou, A.; Karakosta, C.; Mexis, F.D.; Andreoulaki, I.; Psarras, J. A Fuzzy PROMETHEE Method for Evaluating Strategies towards a Cross-Country Renewable Energy Cooperation: The Cases of Egypt and Morocco. *Energies* **2024**, *17*, 4904. https://doi.org/10.3390/en17194904.
58. Castronovo, L.; Filippone, G.; Galici, M.; La Rosa, G.; Tabacchi, M.E. Ontology Aggregation with Maximum Consensus Based on a Fuzzy Multi-criteria Group Decision-Making Method. In Proceedings of the Advances in Fuzzy Logic and Technology; Baczyński, M., De Baets, B., Holčapek, M., Kreinovich, V., Medina, J., Eds.; Springer: Cham, Switzerland, 2025; pp. 76–87. https://doi.org/10.1007/978-3-031-97228-7_7.
59. Gökmen, O.B.; Güven, Y.; Kumbasar, T. FAME: Introducing Fuzzy Additive Models for Explainable AI. In Proceedings of the 2025 IEEE International Conference on Fuzzy Systems (FUZZ), Reims, France, 6–10 July 2025; pp. 1–6. https://doi.org/10.1109/FUZZ62266.2025.11152211.
60. Ruspini, E.H. A new approach to clustering. *Inf. Control* **1969**, *15*, 22–32. https://doi.org/10.1016/S0019-9958(69)90591-9.
61. Wang, L.X.; Mendel, J. Generating fuzzy rules by learning from examples. *IEEE Trans. Syst. ManCybern.* **1992**, *22*, 1414–1427. https://doi.org/10.1109/21.199466.
62. Dunn, J.C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *J. Cybern.* **1973**, *3*, 32–57. https://doi.org/10.1080/01969727308546046.
63. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*, 1st ed.; Advanced Applications in Pattern Recognition; Springer: New York, NY, USA, 1981; p. 272. https://doi.org/10.1007/978-1-4757-0450-1.
64. Hernández-Cano, A.; Hägele, A.; Huang, A.H.; Romanou, A.; Solergibert, A.J.; Pasztor, B.; Messmer, B.; Garbaya, D.; Ďurech, E.F.; Hakimi, I.; et al. Apertus: Democratizing Open and Compliant LLMs for Global Language Environments. *arXiv* **2025**, https://doi.org/10.48550/arXiv.2509.14233.
65. Google DeepMind. Gemma 4 Model Card. 2026. Available online: https://ai.google.dev/gemma/docs/core/model_card_4 (accessed on 7 April 2026).
66. Mistral AI Team. Mistral Small 3. 2025. Available online: https://mistral.ai/news/mistral-small-3 (accessed on 9 March 2026).
67. Yang, Z.; Liu, Z.; Chen, Y.; Dai, W.; Wang, B.; Lin, S.C.; Lee, C.; Chen, Y.; Jiang, D.; He, J.; et al. Nemotron-Cascade 2: Post-Training LLMs with Cascade RL and Multi-Domain On-Policy Distillation. *arXiv* **2026**, https://doi.org/10.48550/arXiv.2603.19220.
68. Kendall, M.G. A new measure of rank correlation. *Biometrika* **1938**, *30*, 81–93. https://doi.org/10.1093/biomet/30.1-2.81.
69. Richardson, M.; Domingos, P. Markov logic networks. *Mach. Learn.* **2006**, *62*, 107–136. https://doi.org/10.1007/s10994-006-5833-1.
70. De Raedt, L.; Kimmig, A.; Toivonen, H. ProbLog: A probabilistic prolog and its application in link discovery. In *IJCAI'07: Proceedings of the 20th International Joint Conference on Artifical Intelligence*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2007; pp. 2468–2473. Available online: https://dl.acm.org/doi/10.5555/1625275.1625673. (accessed on 28 April 2026).