processes

MDPI

Article

Hybrid Fuzzy MCDM for Process-Aware Optimization of Agile Scaling in Industrial Software Projects

Issa Atoum [1,*], Ahmed Ali Otoom [2], Mahmoud Baklizi [3] and Fatimah Alkomah [4]

[1] Department of Software Engineering, Philadelphia University, Amman 19392, Jordan
[2] Department of Cybersecurity and Cloud Computing, Applied Science Private University, Amman 11937, Jordan; a_otoom@asu.edu.jo
[3] Petra University, Amman 11196, Jordan; mbaklizi@uop.edu.jo
[4] Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia; fsalkomah@pnu.edu.sa
* Correspondence: iatoum@philadelphia.edu.jo

Abstract

Scaling Agile in industrial software projects is a process control problem that must balance governance, scalability, and adaptability while keeping decisions auditable. We present a hybrid fuzzy multi-criteria decision-making (MCDM) framework that combines Fuzzy Analytic Hierarchy Process (FAHP) for uncertainty-aware weighting with a tunable VIKOR–PROMETHEE ranking stage. Weighting and ranking are kept distinct to support traceability and parameter sensitivity. A three-layer hierarchy organizes twenty-two criteria across organizational, project, group, and framework levels. In a single-enterprise validation with two independent expert panels (n = 10 practitioners), the tuned hybrid achieved lower rank error than single-method baselines (mean absolute error, MAE = 1.03; Spearman ρ = 0.53) using pre-specified thresholds and a transparent $\alpha + \beta = 1$ control. The procedure is practical for process governance: elicit priorities, derive fuzzy weights, apply the hybrid ranking, and verify stability with sensitivity analysis. The framework operationalizes modeling, optimization, control, and monitoring of scaling decisions, making trade-offs explicit and reproducible in industrial settings.

Keywords: process optimization; fuzzy MCDM; FAHP; VIKOR–PROMETHEE; governance control; industrial software processes; sensitivity analysis; monitoring

1. Introduction

The way an organization scales Agile sets the cadence of planning, allocates decision rights, and governs how feedback flows across distributed teams. When scaling choices are ad hoc or vendor-specific, common outcomes include fragmented practices, duplicated effort, and uneven results [1–3]. The problem is not whether to scale, but how to reach scale while maintaining responsiveness, coherence, and system-level alignment. In industrial settings, these choices should be modeled, optimized, controlled, and monitored with the same discipline applied to production processes. Our aim is practical: make the trade-offs explicit and leave a transparent record of why a choice was made.

Frameworks like SAFe, LeSS, and Disciplined Agile Delivery (DAD) take different positions on governance and coordination [4–6]. Choosing among them is especially challenging in globally distributed or regulated work, where decisions often rest on intuition rather than a structured review [7]. A better approach is to consider scale, governance, and agility within a single development framework and to handle uncertainty explicitly. To this end, fuzzy logic helps by capturing degrees of expert confidence and reducing

Check for updates

Academic Editors: Yuanqiang Zhou, Le Yao, Xiaoyu Jiang and Zheren Zhu

Received: 5 November 2025
Revised: 29 December 2025
Accepted: 3 January 2026
Published: 9 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

ambiguity in socio-technical settings [8]. The practical goal is a workable middle ground: too much control slows adaptation, and too little weakens accountability. Decisions that keep this balance tend to foster maturity, reduce waste, and strengthen resilience over time [9–11]. Recent sector evidence also shows that MCDM (multicriteria decision-making) supports software company strategy and Agile attribute evaluation in operational settings [12,13]. Related work presents a web-based Pythagorean fuzzy decision support system for sustainable project evaluation that combines PF-MEREC (objective weighting) with PF-MARCOS (compromise ranking), which underscores the need for auditable decisions under uncertainty [14]. In software settings, a data-driven model integrates sustainability metrics into Agile project and portfolio assessment, reinforcing governance-aware, auditable choices [15].

1.1. Research Background and Problem Definition

Team priorities are often influenced by three competing dimensions: organizational policy, project constraints, and execution capacity. No single framework optimizes all three. Selection improves when the read is structured and context-aware. FAHP captures graded expert judgments and is common in requirements, maturity, and cost work [16–18]. Yet many applications fuzz the weights while keeping the ranking deterministic; as conditions change, trade-offs can appear fixed when they are not.

Evidence from practice points to contextual fit rather than a single best framework [19]. SAFe emphasizes synchronized increments and a clearer hierarchy; LeSS and Scrum-of-Scrums favor lean, decentralized coordination [4–6]. Each stance adjusts autonomy and oversight differently. Existing decision aids, whether FAHP-based, Cynefin-oriented, or rule-driven, often compress interdependencies among organizational systems, project types, and team dynamics. Misalignment then arises between the framework design and governance intent, leading to inefficiencies and lower stakeholder participation [2,3,7,20–23]. What is still missing is a transparent, reproducible way to support uncertainty-aware weighting and scoring.

1.2. Study Approach and Rationale

We develop and validate an adaptive fuzzy MCDM framework for Agile scaling. Twenty-two criteria, synthesized from the literature and refined with expert input, are organized into four layers: organizational, project, group, and framework. Expert judgments are propagated through fuzzy elicitation and scoring, with clear consistency and propagation checks. For ranking, we integrate VIKOR (compromise) and PROMETHEE (preference flow) to consider agility and control together [24–26]. Fuzziness is used for uncertainty-aware elicitation and propagation during weighting and scoring, while the final ranking is computed on defuzzified (crisp) quantities and audited via sensitivity and agreement metrics. Thus, the approach carries uncertainty forward, maintains auditable steps, and makes trade-offs explicit.

1.3. Research Objectives and Contributions

Our contribution is a reproducible, sustainability-informed framework that helps organizations choose scaling approaches aligned with governance maturity, scalability needs, and sustainability goals. We organize the work around four objectives:

O1. Model a multi-level evaluation hierarchy that carries organizational, project, and framework information through to auditable decisions.
O2. Optimize rankings under uncertainty by combining FAHP weighting with VIKOR–PROMETHEE aggregation and an explicit α–β control.

O3. Control and monitor decision quality using pre-specified accuracy and agreement thresholds (MAE, Spearman ρ) against independent expert panels.
O4. Provide a reproducible, low-friction procedure that organizations can adopt for process-aware scaling decisions, including sensitivity analysis and parameter reporting.

The framework brings dispersed criteria into one layered model with an explicit sustainability mapping, uses triangular fuzzy numbers (TFNs) for uncertainty-aware elicitation and scoring, and computes the final ranking on defuzzified utilities, with robustness assessed through score harmonization.

Section 2 reviews weighting and ranking approaches and related literature; Section 3 details the methodological framework; Section 4 reports empirical findings and expert validation; Sections 5 and 6 conclude with implications, limitations, and directions for further research.

2. Background and Related Work

Software engineering methods run along a spectrum, from the plan-first rigor of Waterfall to the iterative cadence of Agile [27]. At the enterprise level, selecting a scaling framework is a multi-criteria problem: objectives compete, constraints interact, and trade-offs can be missed without a structured approach. Multi-criteria decision-making (MCDM) provides structure by weighting criteria and ranking alternatives, making trade-offs explicit rather than implicit [17]. Prior studies indicate a bidirectional relationship: Agile practices can improve sustainability across the economic, social, and environmental pillars, and sustainability objectives can inform governance and method-selection decisions [28–30].

2.1. Criteria Weighting Methods

Quick diagnostic baselines (e.g., equal weights or single-metric checks) are fast but can obscure decisive criteria [31,32]. Delphi rounds gather expert judgment and build agreement, but they take time and may be influenced by group dynamics [33,34]. The Best–Worst Method narrows the task to the most and least important criteria, reducing effort and yielding steadier judgments [35]. CRITIC and related data-driven options estimate weights from contrast and cross-criterion correlation, providing a solid empirical starting point [36]. In software trustworthiness assessments, FAHP combined with CRITIC has produced stable, audit-friendly weights [37]. Entropy serves as a diagnostic for internal consistency in complex software architectures, offering traceable, information-based validation cues [38]; while entropy weighting is popular, care is needed to ensure it reflects true contrast among criteria [39].

Analytic Hierarchy Process (AHP) and its fuzzy extension (FAHP) are widely adopted. They combine expert judgment with internal consistency checks to make sure that the comparisons hold up [40–42]. In practice, they are used for project scheduling, test-coverage prioritization, and selecting Agile practices [27,43]. FAHP adds a practical layer by mapping linguistic terms (e.g., important, moderate) to triangular fuzzy numbers, preserving nuance in expert opinion. Recent applications of FAHP span vendor evaluation, security assessment, requirements prioritization, and effort estimation [16,18,44,45]. FAHP is also used for sustainable project selection, where fuzzy pairwise judgments operationalize triple-bottom-line criteria in practice [46]. In line with real public procurement cases, trust then depends on making weights transparent and reporting sensitivity [11].

2.2. Ranking Methods

Once the weights are set, the next practical step is to ask how close each candidate is to an ideal. TOPSIS gauges proximity to a positive ideal and separation from a negative one [47–50]. Practitioners use it in Agile assessments, component selection, and sustainabil-

ity analyses [51–53]. Because the distance is Euclidean, structure can compress when many dimensions are in play. VIKOR takes a compromise view, balancing the preferences of most criteria against the most considerable single-criterion regret [24]. When objectives conflict, the compromise helps in practice. Published uses include reliability models, antivirus selection, and software risk studies [54]. Extensions to fuzzy and type-2 fuzzy settings have demonstrated VIKOR's suitability under high uncertainty [55]. Parallel modernization of composite scoring via improved Combined Compromise Solution (CoCoSo) further supports robust aggregation in cloud/service decisions [56].

Evaluation based on Distance from Average Solution (EDAS) uses the average as its anchor and records positive and negative departures for each option [57]. It works well for infrastructure and network plans [58,59], yet a strict focus on the mean can understate qualitative input and expert nuance. PROMETHEE compares options two at a time and applies nonlinear preference functions to model graded favorability [60,61]. With careful parameter tuning, it performs well for software quality assessments and supplier selection [62,63]. Empirical evidence also shows the benefit of hybridizing CRITIC with PROMETHEE for infrastructure and sustainability planning [64]. CoCoSo blends several scoring schemes to stabilize the final ranking, as in cloud-service selection and green-supplier screening [65]. A longitudinal analysis traces PROMETHEE's knowledge evolution and usage patterns across domains [66], and related preference-flow decision-support tools, such as FITradeoff, are increasingly adopted in supplier and governance contexts [67].

2.3. Hybrid MCDM in Software Engineering

This section positions our work within prior hybrid MCDM studies.

2.3.1. Related Work

Hybrid methods connect weighting with ranking so teams can see the trade-offs and trace the reasoning end to end [22]. Comparable fuzzy hybrid MCDM methods have been successfully used in regulated healthcare decisions, confirming their suitability for compliance-intensive settings [68]. These choices sit within a broader Multi-Criteria Decision Analysis (MCDA) landscape, comprehensively surveyed in [69] and, for resilience/supplier decisions, in [67,70]. On the ground, AHP–TOPSIS is reported frequently in portfolio picks and requirement priorities [71,72]; CRITIC–PROMETHEE fits better when subtle preference signals deserve more weight [73]. The same pattern turns up in web service choices, smart city plans, and supplier screens [74–76]. In innovation settings, hierarchical weighting gives stakeholders a shared basis for compromise [10]. The familiar shortcoming is that many setups remain single-layered, overlooking how governance, project constraints, and day-to-day practice push and pull each other.

Recent work on supplier selection applies comparable hybrid methods to trade off lean, agile, resilient, and green objectives [77]. Entropy–VIKOR is used in practice for cybersecurity risk assessment, reliability analysis, and supplier or quality evaluation [54,76,78,79]. Weights that are fixed or driven solely by data tend to lag as project conditions shift. In Agile contexts, AHP–TOPSIS and AHP–PROMETHEE are reported to improve framework selection [80], and fuzzy AHP helps teams prioritize practices [27]. Decision-tree and Grounded Theory studies complement this by providing grounded explanations; however, they scale less easily and rarely include quantitative validation [3,81].

Recent taxonomies place SAFe, LeSS, DAD, and Scrum@Scale along maturity bands [5]. Empirical reviews tell the same story: outcomes hinge on contextual fit rather than on any single best framework [19]. Head-to-head comparisons of SAFe and Scrum of Scrums highlight a clear trade-off: tighter control versus greater flexibility [4]. Operational guidance that is both context-aware and evidence-based remains limited.

2.3.2. Sustainability in Agile and Project Management

We treat sustainability as a first-class decision objective that spans the economic (cost, throughput stability), social (team autonomy, well-being, knowledge continuity), and environmental (travel, compute, facilities) pillars. In Agile and portfolio contexts, prior studies show that these pillars map to measurable indicators used in practice, for example, coordination waste and queueing losses for economic, team autonomy and rework-related burnout proxies for social, and avoided travel miles and build and compute intensity for environmental [1,3,14,15]. This evidence-based perspective positions sustainability as a practical and influential factor in Agile assessment and project selection, including at scale.

2.3.3. Comparison with Existing Agile Scaling Selection Approaches

Evidence shows that context, not brand, drives outcomes in large-scale Agile. Vendor tools (e.g., SAFe self-assessments) work inside their ecosystems and do not compare across families; LeSS guides are pragmatic yet anecdotal; Cynefin frames complexity without yielding a ranked choice [1,2,82]. Recent reviews echo the same point: fit and governance maturity matter more than any single framework [7,19,83].

Within software engineering, MCDM is widely used to make trade-offs explicit. Weighting methods such as AHP/FAHP combine expert judgment with consistency checks, while data-driven options like CRITIC or entropy add contrast where evidence exists [26,36,39–42]. For ranking, established choices include TOPSIS, VIKOR, EDAS, PROMETHEE, and CoCoSo, applied from practice selection to vendor and reliability analyses [24,50,51,56–58,60,65,71,76]. Table 1 lists only approaches that yield a cross-framework ranking, aligning with prior SE evidence [27,37,54,55].

Table 1. Vendor-neutral methods that produce cross-framework rankings.

Method (Family)	Typical Use in SE	Fit to Scaling-Framework Choice	Key Refs.
AHP–TOPSIS (crisp/fuzzy)	Agile practice selection; component/vendor choice	Transparent weights with a clear closeness score across alternatives	[27,40,41,50,51,71]
CRITIC–PROMETHEE	Software quality and supplier ranking; infrastructure trade-offs	Good when graded preference functions matter; supports sensitivity analysis	[36,60,64,73]
Entropy–VIKOR (type-1/2 fuzzy)	Reliability and cyber-risk; supplier selection	Robust under conflicting goals; makes the compromise option explicit	[24,54,55,76]
EDAS	Network and infrastructure evaluation	Simple, explainable scoring; useful as a robustness check with TOPSIS/VIKOR	[57,58]
CoCoSo (improved variants)	Cloud/service provider decisions	Stabilizes ranks when criteria pull differently; triangulates with the above	[56,65]

Note: "Fit" summarizes why each method works for cross-framework selection rather than within one vendor ecosystem.

Compared with these strands, our study emphasizes explicit treatment of uncertainty, end-to-end traceability from weighting to ranking, and first-class sustainability criteria, consistent with contingency-fit perspectives and systems thinking in large-scale change [1,14,15,29,30,84].

2.3.4. Identified Research Gaps and Study Positioning

Despite steady progress, four gaps remain. First, two-layer structures help organize decisions but rarely trace how uncertainty moves or how feedback travels across levels [85]. Second, many fuzzy compromise-ranking methods improve resilience assessments but seldom fix inter-criterion dynamics [70]. Third, fuzzy predictive systems handle uncertainty well, but they typically optimize a single objective rather than a balanced, multi-level

set [86]. Moreover, despite PROMETHEE's methodological maturation [66], many implementations still under-document preference-function choices and sensitivity ranges, limiting auditability in organizational settings. In parallel, machine learning decision tools assist design, forecasting, and resource planning [87–89], while efficiency improves, transparency and auditability often do not. We introduce a fuzzy MCDM pipeline that uses TFNs for uncertainty-aware weighting and scoring, then computes the final ranking on defuzzified utilities.

We exclude vendor-anchored, anecdotal, or non-quantitative sources because they cannot deliver a reproducible, auditable cross-framework ranking with traceable weights and sensitivity. In contrast, the methods in Table 1 do, and they align with established SE uses of MCDM [82,90,91].

3. Framework Ranking and Tunable Hybrid Control

We frame scaling-framework selection as a five-stage process of criteria screening, weighting, framework ranking, framework selection, and validation (Figure 1). The central part is a hierarchical decision structure that elicits weights from organizational and project factors, then applies a shared evaluation rubric to score framework groups, with full operational definitions provided in Supplementary Sections S3.1–S3.5 [6,73,92].

Hierarchy clarification (O, P, G, C). To avoid ambiguity in terminology, we distinguish between weight-elicitation layers and evaluation layers. Organizational factors ($O1$–$O7$) and project factors ($P1$–$P6$) capture context and priorities and are propagated through fuzzy AHP. Candidate scaling approaches are represented as seven framework groups ($G1$–$G7$) that form the alternative set to be ranked. Each group is evaluated using a shared measurement rubric ($C1$–$C9$) that operationalizes scoring and comparison, where $C1$–$C9$ functions as a rubric rather than an additional decision tier. Detailed definitions, sources, and expert-panel validation evidence are provided in Supplementary Sections S3.1–S3.5, with a SAFe/LeSS/DAD construct crosswalk in the Supplementary Material.

Panel alignment and context. Before detailing Stages 1–5, we clarify how the expert panels align with the workflow in Figure 1. Both panels came from the same large technology enterprise, so weighting and ranking reflect one coherent context. Panel A included strategic leaders (PMO, portfolio, governance) who refined and weighted the organizational, project, and sustainability criteria via Delphi and fuzzy AHP. Panel B comprised delivery managers and product leaders who then applied these weights to rank candidate scaling frameworks for real projects. To keep evaluations concrete, we organized the portfolio into seven project cohorts (PC1–PC7), each grouping two to three comparable projects at similar maturity levels. This design anchors the reported rankings in actual projects rather than hypothetical cases, linking enterprise priorities from Panel A with project-level evaluations from Panel B.

3.1. Stage 1: Criteria Screening

Stage 1 sets a simple three-level structure: organizational, project, and group. We used GPT-4 as a light co-pilot to list, cluster, and de-duplicate candidate criteria from the literature and our experience. We seeded it with the core scaling themes (governance, maturity, flow, risk, adaptability) and asked it to consolidate overlapping prompts before expert review [93].

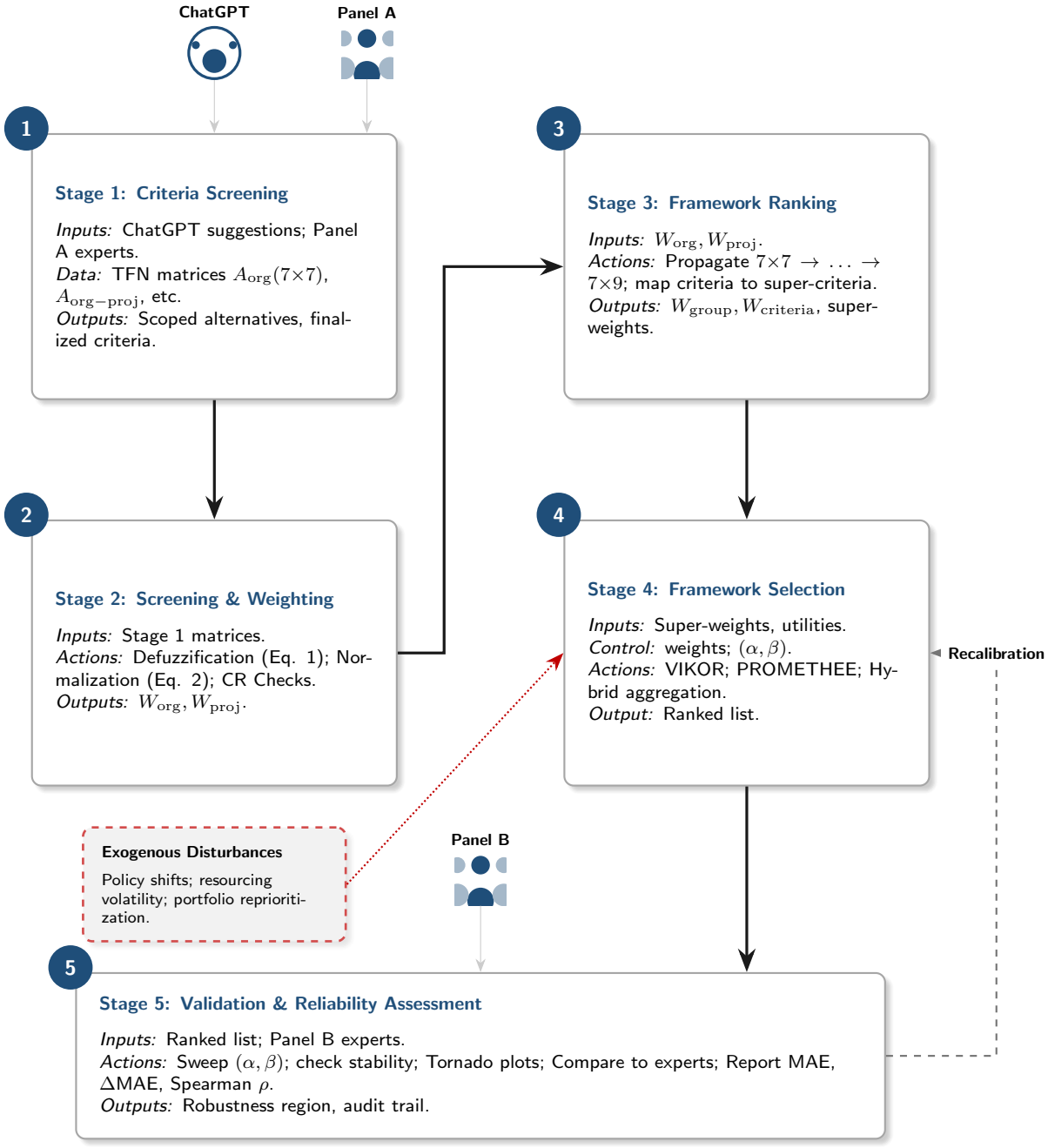


Figure 1. Governance-aware decision workflow. Stage 1 screens and structures criteria; Stages 2–4 propagate weights and compute rankings under tunable (α, β) settings; Stage 5 validates with a separate expert panel using monitoring indicators (agreement, error, stability). The workflow also includes an exogenous-disturbance note and a recalibration loop to update settings when stability degrades.

Large language models help draft the pool of candidate criteria and examples [94–96]. This careful, value-based use of GPT-4 matches recent guidance for software organizations [97]. We followed disciplined prompt-pattern practices to bound GPT-4o's role strictly to phrasing and clustering support [98]. We tried several prompt versions and different criteria counts, then settled on 22 items because this number fits human cognition in workshops, keeps coverage across the three levels, and remains fine-grained yet realistic. We stage criteria by level so experts avoid scoring everything at once and can tailor or prune items as needed. Sustainability is treated as a cross-cutting property within the existing

22 criteria to keep the hierarchy compact. Economic and social sustainability are directly captured via cost, time-to-value, autonomy, and collaboration criteria, while the environmental pillar is represented indirectly through dispersion and waste-reduction proxies (e.g., P4 and lean/quality-related criteria), with direct energy or travel measures left as an extension when such data are available. The complete set of 22 candidates appears in the Supplementary File. For example, organizational criteria cover governance and Agile maturity to gauge oversight and adoption depth; project criteria focus on requirements volatility and team distribution to capture change and coordination, while group criteria emphasize scaling frameworks and lean/flow to align portfolio coordination and waste reduction.

The 22 criteria were clustered with limited LLM assistance, then finalized by Panel A through two Delphi rounds. Panel A supplied all TFN pairwise and link matrices ($A_{\text{org}} \in \mathbb{R}^{7\times7\times3}$, $A_{\text{org–proj}} \in \mathbb{R}^{7\times6\times3}$, $A_{\text{proj–group}} \in \mathbb{R}^{6\times7\times3}$, $A_{\text{group–criteria}} \in \mathbb{R}^{7\times9\times3}$); weights were centroid-defuzzified, column-normalized, propagated, and checked for CR $<$ 0.10. Saaty's CR is computed only for reciprocal square pairwise-comparison matrices and is not defined for rectangular cross-level link matrices. For link matrices, we report dispersion and convergence diagnostics on defuzzified entries to identify unstable mappings; see Table S12 for the full ledger.

3.2. Stage 2: Multi-Level Screening and Weighting

All weighting in the proposed approach relies on triangular fuzzy numbers (TFNs) and is defuzzified with centroid values (Equation (1)). TFNs were chosen to minimize elicitation burden while keeping uncertainty representation transparent for audit; rank stability was checked under small perturbations of TFN spreads, while centroid-matched trapezoidal and type-2 variants are left for future work. Matrices are column-normalized (Equation (2)) and propagated through the hierarchy via matrix multiplication (Equation (3)). Keeping these steps distinct makes the reasoning auditable and curbs bias. At the organizational layer, a $7 \times 7 \times 3$ TFN matrix encodes strategic priorities such as governance, scaling capacity, delivery cadence, and cost efficiency. The weights map to the project layer as a $7 \times 6 \times 3$ structure covering application type, requirements complexity, team distribution, stakeholder engagement, and time-to-value. Project-level weights then feed into a $6 \times 7 \times 3$ representation of framework-adoption groups (Scaling, Lean/Flow, Team-Centric, Governance, Hybrid/Mixed, Risk-Oriented, Continuous Delivery). Next, group criteria are consolidated into three higher-order decision axes, Capability Alignment (CA), Scaling Fitness (SF), and Governance and Adaptability Maturity (GAM), via the mapping $\widehat{A}_{\text{criteria–super}}$ (Equation (4)). The mapping matrix $\widehat{A}_{\text{criteria–super}} \in \mathbb{R}^{9\times3}$ is centroid-defuzzified and column-normalized (Equation (2)) so each super-criterion column sums to 1, yielding a comparable CA/SF/GAM contribution scale. This yields super-criteria weights that reflect dominant priorities in the organizational context. The super-criteria then guide the fuzzy screening and subsequent ranking. For clarity, $A_{\text{group–criteria}} \in \mathbb{R}^{7\times9\times3}$ encodes how the seven framework groups ($G1$–$G7$) are scored against the nine rubric criteria ($C1$–$C9$), consistent with Supplementary Sections S3.1–S3.5.

As the ultimate goal is to rank development frameworks, an expert utilizes a $7 \times 9 \times 3$ matrix to evaluate the seven Agile framework groups on nine shared criteria. This produces a TFN score set $\tilde{G}$ with one fuzzy number per framework–dimension pair. Per-group fuzzy scoring (Equation (5)) followed by centroid defuzzification (Equation (1)) yields a crisp decision matrix $X = x_f \in \mathbb{R}^{|F|\times3}$, where $x_f = (s_f^{\text{CA}}, s_f^{\text{SF}}, s_f^{\text{GAM}})$. Optional single-score utilities u_f (Equation (6)) provide concise comparative baselines for subsequent ranking and sensitivity analysis.

After TFN-based elicitation and scoring, centroid defuzzification (Equation (1)) yields the crisp decision matrix X (Equation (5)) and crisp weight vectors (e.g., W_{super}); all subsequent utilities and ranks are computed on these defuzzified quantities (Equation (7)).

3.3. Stage 3: Framework Ranking Through the Hybrid Method

The crisp decision matrix X produced in Stage 2 after centroid defuzzification (Equation (1)) feeds directly into Stage 3, where criteria weights $\omega \in$ FAHP, Entropy, CRITIC are applied within each ranking engine to compute method-specific utilities (Equation (7)).

3.3.1. Weighting and Method-Specific Ranking

Three weighting schemes are considered for ranking: FAHP (defuzzified, from Equation (4)), Entropy, and CRITIC. Each scheme provides a weight vector ω over CA, SF, GAM, applied to the Stage 2 decision matrix $X = x_f$ to compute method-specific utilities $R_m^{(\omega)}(f)$ (Equation (7)). Recent software-oriented studies report that FAHP–CRITIC pairings yield explainable yet contrast-sensitive weights, while entropy can serve as a useful baseline when empirical dispersion is meaningful [37,39].

On the weighted matrix, five ranking methods are applied: TOPSIS (proximity to the ideal solution), VIKOR (compromise-oriented), PROMETHEE (preference modeling), EDAS (distance from the average), and M-TOPSIS (modified robustness). Each method m returns a utility score $R_m^{(\omega)}(f)$ for every framework f (Equation (7)); an optional single-score utility u_f (Equation (6)) is also computed for baseline comparison. In each variant, ω serves as the criteria weight vector on CA, SF, GAM, replacing W_{super} when evaluating $R_m^{(\omega)}(f)$.

3.3.2. Hybrid Aggregation of Rankings

Because no single method captures all decision perspectives, selected outputs are combined using the hybrid function (Equation (8)). Tunable parameters α and β with $\alpha + \beta = 1$ control relative contributions, for example, balancing VIKOR's compromise logic against PROMETHEE's preference-flow modeling. Validation metrics are embedded within this aggregation: mean absolute error (MAE) quantifies deviation from expert-provided ranks, and Spearman's correlation (ρ) measures convergence across methods.

3.3.3. Sustainability Signals Within Process Criteria (CA, SF, GAM)

We integrate the triple bottom line within the existing super-criteria rather than adding a new axis: economic (C9 cost efficiency, C6 implementation time and effort, P6 time to value), social (C7 team autonomy, C8 cross-functional collaboration, P5 stakeholder engagement), and environmental (P4 team distribution as a proxy for travel and digital resource use). These factors flow into CA, SF, and GAM through the 9×3 mapping in Equation (4): CA captures size and complexity fit, SF captures adaptability, and GAM captures oversight and organizational maturity. Emphasis can be tuned by editing rows of $A_{\text{criteria-super}}$ and the (α, β) mix (Equation (8)); performance is evidenced by the reported ρ and MAE.

3.4. Stage 4: Framework Selection and α–β Tuning

Stage 4 turns the hybrid scores into a calibrated recommendation. We tune the balance between algorithms using α–β control $\alpha \in \{0.1, 0.2, \ldots, 0.9\}$ with $\beta = 1 - \alpha$, score each pair with mean absolute error against expert ranks, and track agreement using Spearman's ρ. The target is the α–β setting that maximizes agreement while minimizing deviation, shifting from a fixed blend to an empirically calibrated balance.

The intuition is that context matters. Regulated domains (finance, healthcare) can weigh governance, while start-ups or product teams can favor delivery speed. The tuned ranking remains transparent and adjustable: stakeholders see how priorities influence outcomes.

3.5. Stage 5: Validation and Reliability Assessment

Stage 5 asks whether the tuned ranking holds up against experienced practitioners and whether the route to that ranking is auditable. We use two independent panels and keep automation in a supporting role only.

Panel setup and elicitation. Panel A (criteria) conducts Delphi-style iterations to refine definitions and produce fuzzy pairwise comparisons for weighting. Panel B (frameworks) applies the finalized rubric to score each alternative on CA, SF, and GAM. Roles do not overlap. As documented in Supplementary Table S12, both panels were purposively sampled (n = 5 each) with role separation (weighting vs. scoring). Panel B was blinded to Panel A outputs and submitted scores independently; bias controls included structured elicitation, anonymized aggregation, and conflict-of-interest screening.

Planned analysis. We compare model ranks to expert ranks using mean absolute error (MAE) and Spearman's ρ (Equation (9)); when a group has only two items, ρ is omitted. We report MAE in rank-position units (e.g., MAE $= 1$ means an average deviation of one rank position between the model order and the expert order). Rank error is the absolute rank-position deviation $|r_{\text{agg}}(f) - r_{\text{exp}}(f)|$, MAE is its average across alternatives, and Spearman's ρ is computed on complete rank vectors (Equation (9)). In the industrial tuning step, we use these metrics to calibrate α–β by minimizing rank displacement and maximizing rank-order stability, and we report the sensitivity surface rather than asserting universal cutoffs. Inter-rater reliability is assessed with two-way random-effects intraclass correlation, ICC(2,1), for Panel A's defuzzified weights and Panel B's per-dimension scores.

Baseline Construction For each framework group g, we compute the baseline MAE as the average of PROMETHEE, TOPSIS, and VIKOR MAEs within that group. The overall baseline is the group-size weighted average.

Acceptance rule. We accept a tuned α–β configuration when MAE falls within predefined tolerance bands and ρ indicates at least moderate agreement, with the full audit trail (rationales and dispersion summaries) available for managerial review.

3.6. Algorithmic Workflow Overview

The computational workflow proceeds in two phases. For quick reference, Equations (1)–(11) are collected in Table 2 along with their notation.

Phase 1: Fuzzy Weighting and Propagation. Expert pairwise comparisons are encoded as triangular fuzzy numbers and defuzzified by the centroid (Equation (1)); columns are normalized (Equation (2)). We propagate weights through the hierarchy (Equation (3)) and map nine criteria to three super-criteria (Equation (4)), embedding the sustainability mapping described above.

Phase 2: Hybrid Ranking and Validation. Framework scores on CA, SF, and GAM are normalized within groups and defuzzified (Equation (5)). Multiple rankers (TOPSIS, VIKOR, PROMETHEE, EDAS, M-TOPSIS) run under FAHP/Entropy/CRITIC weights (Equation (7)). A tunable hybrid (Equation (8)) combines complementary logics with $\alpha + \beta = 1$. Agreement with expert panels is assessed via MAE and Spearman's ρ (Equation (9)). Detailed pseudo-code is provided in the Supplementary Materials.

All pairwise judgments and framework scores come from human experts via anonymous rounds; large language models only helped draft candidate criteria and wording. We log short rationales with each judgment and archive inputs, weights, and intermediate scores so that anybody can rerun the pipeline and trace every rank end-to-end.

Table 2. Core equations and notation used throughout the study.

(1) Centroid Defuzzification:	$\hat{a}_{ij} = \frac{1}{3}(l_{ij} + m_{ij} + u_{ij})$	(1)
(2) Column Normalization:	$\hat{a}'_{ij} = \frac{\hat{a}_{ij}}{\sum_{i=1}^{n} \hat{a}_{ij}}$	(2)
(3) Hierarchical Propagation (to 9 criteria):	$W_{\text{criteria}} = W_{\text{org}}\, \widehat{A}_{\text{org–proj}}\, \widehat{A}_{\text{proj–group}}\, \widehat{A}_{\text{group–criteria}}$	(3)
(4) Criteria → Super-Criteria (9×3):	$W_{\text{super}} = W_{\text{criteria}}\, \widehat{A}_{\text{criteria–super}}$	(4)
(5) Per-Group Fuzzy Scoring:	$\tilde{S}_{f,c} = \text{norm}_{g(f)}\big(\tilde{G}_{f,c} \odot \tilde{W}_{g(f)}\big),$ $x_f = \big(m(\tilde{S}_{f,\text{CA}}),\ m(\tilde{S}_{f,\text{SF}}),\ m(\tilde{S}_{f,\text{GAM}})\big)$	(5)
(6) Framework Utility (baseline):	$u_f = (w_{\text{CA}}, w_{\text{SF}}, w_{\text{GAM}}) \cdot \begin{bmatrix} s_f^{\text{CA}} \\ s_f^{\text{SF}} \\ s_f^{\text{GAM}} \end{bmatrix},$ with $(w_{\text{CA}}, w_{\text{SF}}, w_{\text{GAM}}) = W_{\text{super}}$	(6)
(7) Method-Specific Utility:	$R_m^{(\omega)}(f) = \Psi_m(X, \omega),$ $X = \{x_f\} \in \mathbb{R}^{\lvert F\rvert \times 3},$ $\omega \in \{W_{\text{super}}\ (\text{FAHP}),\ \text{Entropy},\ \text{CRITIC}\}$	(7)
(8) Hybrid Aggregation:	$\widehat{R}_m(f) = \frac{R_m^{\uparrow}(f) - \min_{f'} R_m^{\uparrow}(f')}{\max_{f'} R_m^{\uparrow}(f') - \min_{f'} R_m^{\uparrow}(f') + \epsilon},$ $R_{\text{agg}}(f) = \alpha\, \widehat{R}_{m_1}^{(\omega_1)}(f) + \beta\, \widehat{R}_{m_2}^{(\omega_2)}(f), \qquad \alpha + \beta = 1.$	(8)
(9) Validation Metrics:	$\rho = \text{Sp}(r_{\text{exp}}, r_{\text{agg}}) = 1 - \frac{6\sum_{f \in F}\big(r_{\text{agg}}(f) - r_{\text{exp}}(f)\big)^2}{\lvert F\rvert\big(\lvert F\rvert^2 - 1\big)},$ $\text{MAE} = \frac{1}{\lvert F\rvert}\sum_{f \in F}\lvert r_{\text{agg}}(f) - r_{\text{exp}}(f)\rvert.$	(9)
(10) Consistency Check (Saaty):	$\lambda_{\max}(A) = \max \text{eig}(A),\ \text{CI} = \frac{\lambda_{\max}(A) - n}{n-1},\ \text{CR} = \frac{\text{CI}}{\text{RI}_n},\quad A \in \mathbb{R}^{n \times n}.$	(10)
(11) Relative Improvement in MAE:	$\Delta\text{MAE}(\%) = \frac{\text{MAE}_{\text{baseline}} - \text{MAE}_{\text{method}}}{\text{MAE}_{\text{baseline}}} \times 100\%$	(11)

Notation. (i) $\odot$ denotes elementwise TFN multiplication; $m(\cdot)$ is the centroid. (ii) Each $\widehat{A}$ is first defuzzified and then column-normalized (Equations (1) and (2)). (iii) Equation (3) maps $7 \times 7 \to 7 \times 6 \to 6 \times 7 \to 7 \times 9$. (iv) Equation (4) introduces a 9×3 link to CA, SF, and GAM. (v) α and β are tunable, with baseline $(0.5, 0.5)$. (vi) CR < 0.10 is treated as acceptable. (vii) Equation (11): % MAE reduction vs. stated baseline (group-size–weighted mean of single-method MAEs). (viii) $R_m^{\uparrow}$ denotes direction-aligned utilities (all transformed to "higher is better"); $\widehat{R}$ is the min–max scaled score in $[0, 1]$.

3.7. Ranking Method Parameterization

We keep settings simple and standard so results are easy to compare and audit. For PROMETHEE, we use the linear preference function (Type I) with no thresholds ($q = p = 0$). We weight pairwise preferences, cut negatives to zero, take the net flow as the utility, and rescale to $[0, 1]$. For *VIKOR*, we use the classic compromise setup with $v = 0.5$ and the standard *S–R–Q* sequence. Preference-function choices can change flows, so we state them for traceability [66]. For *EDAS*, we follow the original procedure: min–max normalization and deviations from the mean, with no extra parameters. For *TOPSIS* and *M-TOPSIS*, we

apply vector normalization and Euclidean distance to the positive and negative ideals; there is no tuning beyond normalization and weights.

All ranking methods share the same weighting schemes, $\omega \in \{FAHP, \text{Entropy}, \text{CRITIC}\}$, as in Equation (7). We considered crisp AHP as a qualitative baseline for traceability rather than as a competing ranker. Before ranking, we normalize inputs within each group (Equation (5)) and rescale utilities to $[0, 1]$ if needed (Equation (7)). Final scores use the α–β hybrid aggregation (Equation (8)); the AHP baseline provides a non-fuzzy reference in Section 4.

3.8. Research Hypotheses

We derive three testable hypotheses from these gaps and the hybrid design:

H1. *Group-based classification improves alignment. Placing frameworks into seven categories (Scaling, Lean/Flow, Team-Centric, Governance, Hybrid/Mixed, Risk-Oriented, Continuous Delivery) will align selections more closely with organizational and project needs than evaluating each framework in isolation.*

H2. *Hybrid ranking improves agreement with experts. A unified hybrid of multiple rankers will show higher Spearman's ρ and lower MAE against expert judgments than any single method.*

H3. *Dynamic α–β weighting enhances stability. Tuning α–β will reduce ranking variance across methods and weighting schemes. We assess stability via a sensitivity sweep over $\alpha \in [0.3, 0.7]$ using Spearman's ρ and MAE.*

Together, these hypotheses target alignment, agreement, and stability; we evaluate each against preset thresholds and comparative baselines.

4. Results

The framework addresses multi-level decisions by turning the method in Section 3 into actionable steps. We present two principal outcomes. First, group-level alignment summarizes how clusters of Agile frameworks perform across nine shared criteria mapped to three super criteria: CA, SF, and GAM. Second, framework level rankings report the final hybrid scores for the complete set of thirty frameworks (listed in Supplementary Materials). For traceability, the Results are reported in the same Stage 1–5 sequence as Section 3. Full intermediate matrices, sensitivity surfaces, and extended figures are provided in the Supplementary Materials (Sections S3.1–S3.5).

Model performance is evaluated against expert judgments using MAE and Spearman's correlation (ρ). To examine robustness, we conduct sensitivity analysis over varying α–β configurations; this helps indicate whether observed patterns are stable rather than artifacts of a single parameter choice. All detailed matrices, weighting propagation steps, and supplementary figures are provided in the Supplementary Materials to support transparency, replicability, and open science practices. Although the dominant trend favors methods that balance governance with delivery speed, one cannot entirely dismiss the critique that rank stability may depend on the salience of local criteria; the sensitivity analysis is intended to surface such effects.

4.1. Results of Stage 1: Criteria Screening

First, group-level alignment shows how clusters of frameworks perform on nine shared criteria mapped to three super-criteria (CA, SF, GAM). Second, framework-level rankings report the final hybrid scores for thirty frameworks (listed in Supplementary Materials). We evaluate performance against expert judgments using MAE and Spearman's ρ, and we

probe robustness with α–β sensitivity. Full matrices are in the Supplementary Materials for transparency and reuse.

4.1.1. Organization-Specific Criteria (O1–O7)

Seven organizational drivers are elicited on a compact fuzzy 1–3 scale: Scaling the organization, Speed of delivery is important, Quality control, Governance and control, Agile maturity, Being flexible and able to change, and Cost effectiveness. In practice, Scaled Agile Framework (SAFe)or Disciplined Agile Delivery (DAD) may rate high on scaling and governance, whereas XP (Extreme Programming) or Crystal may rate low on those dimensions while showing strength on flexibility and speed; Scrum and Enterprise Kanban are typically high on cost effectiveness. The TFN pairwise matrix is defuzzified and column-normalized (Equations (1) and (2)) to yield W_{org}.

4.1.2. Project-Specific Criteria (P1–P6)

Organizational priorities are routed through six project descriptors: type of application (noncritical to critical), size of the application (small to enterprise), requirements volatility and interdependency, team topology and distribution (co-located to fully distributed), Stakeholder involvement (minimal to continuous), and time-to-value (long to rapid). For example, strong governance at the organizational level may increase influence on complexity handling, whereas a high-speed priority naturally maps to delivering value quickly. This produces $W_{\text{proj}} = W_{\text{org}} \widehat{A}_{\text{org–proj}}$ (Equation (3)).

4.1.3. Framework Groups (G1–G7)

Frameworks are organized into seven intent-based clusters: scaling frameworks (G1), lean and flow (G2), team-centric Agile (G3), governance-oriented frameworks (G4), hybrid and mixed-method (G5), risk-driven frameworks (G6), and continuous delivery (G7). Larger or more distributed projects tend to push weight toward G1 and G4; lighter, speed-oriented contexts emphasize G2 and G7.

4.1.4. Criteria for Groups (C1–C9)

Each group is profiled on nine shared criteria: organizational size suitability, complexity handling, governance and control, flexibility and adaptability, Agile maturity requirement, implementation time and effort, team autonomy and empowerment, cross-functional collaboration, and cost efficiency. Expert uncertainty is retained using TFNs rather than collapsing to single scores. For illustration, G1 (scaling frameworks) often scores higher on governance and control (C3) and cross-functional collaboration (C8) than on cost efficiency (C9). Propagation yields $W_{\text{criteria}} = W_{\text{group}} \widehat{A}_{\text{group–criteria}}$.

4.2. Results of Stage 2: Multi-Level Screening

Stage 2 maps how framework groups and individual frameworks align with the Stage 1 organizational and project priorities. We run two passes: a group screening that tests group suitability against shared criteria, and a framework screening that scores individual frameworks within groups across three dimensions. The outputs form a structured decision matrix used as input to the Stage 3 hybrid ranking.

4.2.1. Group Screening Results

Groups are evaluated with a $7 \times 9 \times 3$ matrix against nine shared criteria derived from the organizational and project weights. Expert fuzzy judgements are aggregated and defuzzified to produce coherent group-level profiles. The seven clusters (G1–G7) present distinct, reusable profiles across C1–C9 that we use for screening.

4.2.2. Framework Screening Results

The nine criteria are aligned with three screening axes: Core Alignment (CA), Scalability & Flexibility (SF), and Governance & Agile Maturity (GAM). As guidance, Governance and control (C3) loads primarily on GAM, Flexibility and adaptability (C4) on SF, and Organizational size suitability (C1) together with Complexity handling (C2) contribute to CA. After defuzzification and column-normalization of the 9×3 mapping, we obtain the normalized vector W_{super} aligned to {CA, SF, GAM} (Equation (4)). We note that column-normalization can attenuate minority criteria contributions; we therefore report a sensitivity check against alternative normalizations in the Supplementary (Table S12). This vector is used in Stage 3 to weight method utilities (Equation (6) and in the hybrid aggregation (Equation (8)). The results show a precise match between group profiles and individual performance. In organizations with very tight delivery constraints, the SF dimension can outweigh GAM. Full numeric details are provided in the Supplementary Materials.

4.3. Results of Stage 3: Hybrid Framework Ranking

Stage 3 takes the weighted matrix X, ranks the options with several MCDM methods across different weighting schemes. To keep perspective and comparability, we run five methods across three weighting schemes: TOPSIS, VIKOR, PROMETHEE, EDAS, and M-TOPSIS with FAHP, Entropy, and CRITIC. Each combination returns a utility $R_m^{(\omega)}(f)$ for framework f (Equation (7)). The perspectives are complementary: TOPSIS gauges proximity to an ideal, VIKOR balances utility and regret, PROMETHEE models pairwise preferences, EDAS measures deviation from the mean, and M-TOPSIS adds robustness to outliers. Single-score utilities u_f (Equation (6)) provide a common baseline for comparing the multidimensional rankings.

4.3.1. Hybrid Aggregation Results

Two engines are integrated to mitigate reliance on any single logic with the hybrid function (Equation (8)). Parameters α and β with $\alpha + \beta = 1$ set their relative influence, allowing TOPSIS's proximity view to be balanced against VIKOR's compromise view. As shown in Figure 2, tuning improves agreement with expert evaluations measurably. Across most groups, hybrid ranks achieve MAEs below 1.5 and Spearman correlations above 0.50, outperforming all per-group single-method baselines (PROMETHEE/TOPSIS/VIKOR) and group-size–weighted baselines.

4.3.2. Group Level Observations

Performance varies by group and mirrors differences among candidates. Team-centric (G3) and hybrid/mixed (G5) categories show the strongest agreement with experts (e.g., $\rho \approx 0.67$), followed by scaling (G1; $\rho \approx 0.62$). Governance (G4) and risk-oriented (G6) groups yield moderate correlations (about 0.30–0.33), consistent with overlapping features. Continuous delivery (G7) includes two items, so ρ might not be interpretable; we report MAE instead. Overall, the hybrid approach performs best when alternatives differ meaningfully in emphasis, while remaining interpretable under broader heterogeneity.

4.4. Results of Stage 4: Framework Selection

Stage 4 refines the hybrid ranks by tuning α–β to maximize agreement with expert judgments while reflecting priorities such as governance, adaptability, or compliance. The discrete sweep is reported in Table 3; the $\alpha = 0.3$, $\beta = 0.7$ setting yields the lowest MAE with the highest ρ, matching the narrative in Figure 2.

Table 3. Continuous sensitivity analysis of α–β parameters (Best values are in bold).

α	β	MAE	ρ	Organizational Context
0.1	0.9	1.34	0.41	Startups, high flexibility
0.3	0.7	**1.03**	**0.53**	Balanced organizations
0.5	0.5	1.25	0.45	Transitional contexts
0.7	0.3	1.42	0.39	Regulated industries
0.9	0.1	1.58	0.32	High-compliance domains

Figure 2 shows the trade-off surface: lower MAE is better; higher Spearman's ρ is better. The tuned hybrids improve both metrics over all single-method baselines. Correlations are highest in team-centric and hybrid/mixed groups, moderate in governance- and risk-oriented groups, and MAE is lowest when alternatives are clearly differentiated.

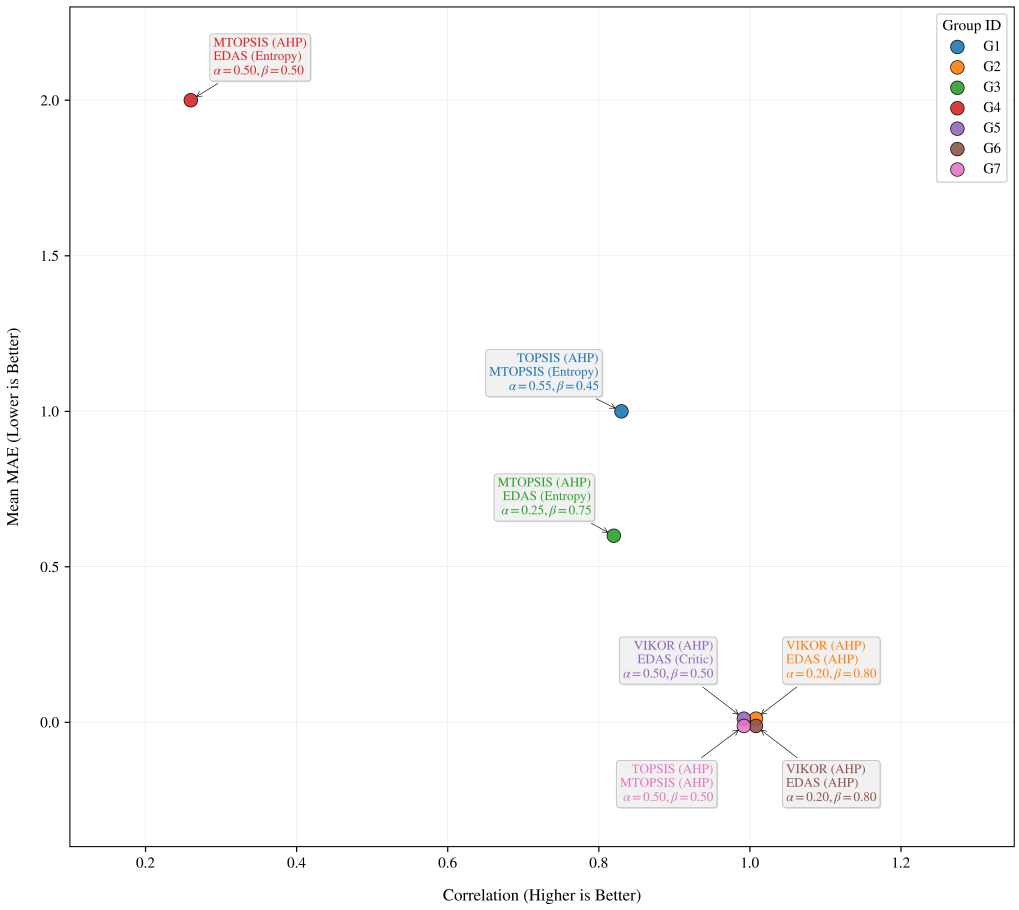


Figure 2. Stage 4: Quality metric space by framework group. Points represent each group's best hybrid configuration, annotated with the method pair and α–β weights. The x-axis is the correlation with expert rankings (higher better); the y-axis is mean absolute error (lower better). Lower-right clusters signal optimal agreement with experts. VIKOR–PROMETHEE consistently gave the steadiest results and the closest match to expert orderings.

Figure 3 shows a simple pattern. Hybrid pairs like TOPSIS–PROMETHEE and VIKOR–PROMETHEE keep scores tight, a good signal of stability. Combinations that include M-TOPSIS move around more when parameters change. The VIKOR–PROMETHEE pair offers the best trade-off at $\alpha = 0.3$ and $\beta = 0.7$, most visible in G2, and it holds steady under the sensitivity checks, as depicted in Figure 4.

Contextual Interpretation

Empirical tuning produces stable ranks. In our validation, governance-oriented frameworks rank highest in regulated domains, while team-centric and lean approaches lead in faster-moving settings. In our single-enterprise validation, the hybrid approach yielded stable recommendations; external contexts may differ.

4.5. Results of Stage 5: Expert Validation

Panel B applied Panel A's defuzzified weights to rank frameworks for the same enterprise's project cohorts (PC1–PC7), each covering 2–3 comparable projects. Panel B scored frameworks on CA, SF, and GAM using the finalized rubric (Tables 4 and 5). Separating weighting from scoring reduced bias and improved reliability.

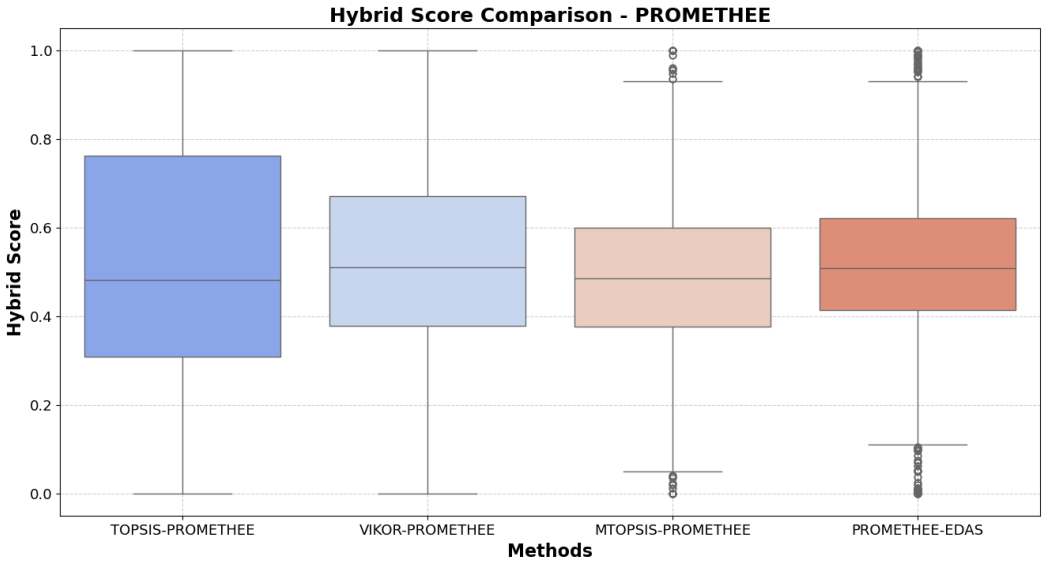


Figure 3. Stage 4: VIKOR–PROMETHEE under Entropy weights. The boxplots show how hybrid scores change as α and β shift. The choice $\alpha = 0.3$, $\beta = 0.7$ offers the most even trade-off, pairing lower MAE with higher ρ.

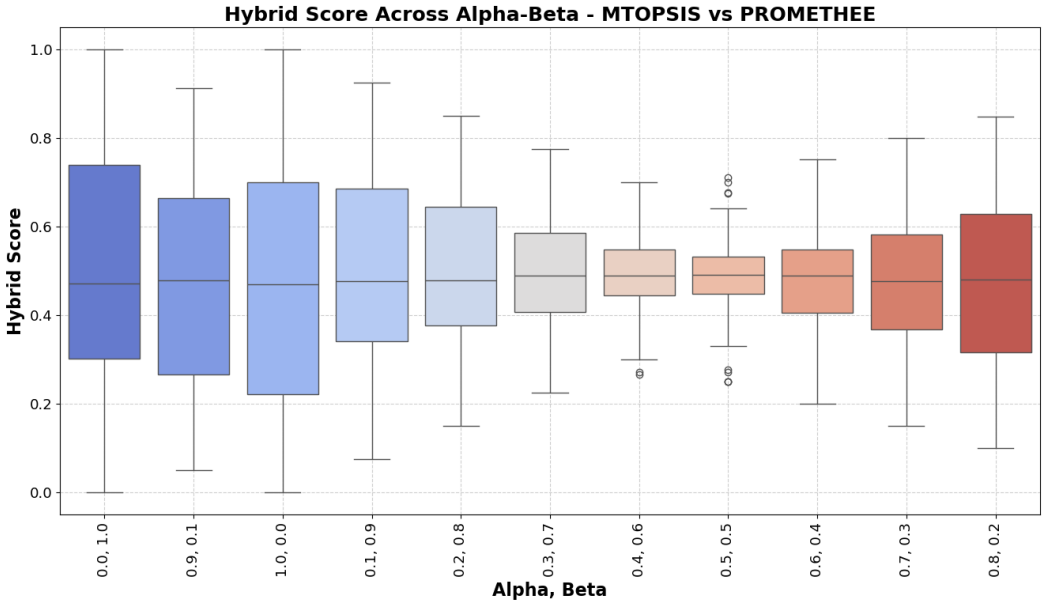


Figure 4. Stage 4: VIKOR–PROMETHEE under Entropy weights, sensitivity across α–β settings. Even as the parameters shift, the hybrid score distributions stay tight, a practical sign of robust behavior.

Table 4. Stages 1–2: Panel A (same enterprise), criteria refinement and weighting.

ID	Role	Experience (Years)	Enterprise Function	Specialization	Panel Scope
A1	Senior Agile Coach	18	PMO, Portfolio Governance	SAFe, LeSS, Prioritization	Org, Project, CA/SF/GAM map
A2	Dev Lead	14	Delivery Engineering	Scrum, XP, DAD	Project criteria, checks
A3	Enterprise Architect	16	Architecture, Standards	Agile Governance, SAFe	Org criteria, propagation
A4	Transformation Consultant	12	Change Enablement	SAFe, Lean, Kanban	Org and Project levels
A5	Agile Coach (CoE)	20	Center of Excellence	LeSS, Scrum, Scaling	Criteria–to–CA/SF/GAM

Table 5. Stage 5: Panel B (same enterprise), framework rankings by project cohorts (PC1–PC7).

ID	Role	Experience (Years)	Program Area	Specialization	Project Cohort(s)
B1	Agile Coach	15	Core Platforms	Scrum, SAFe	PC1–PC2
B2	Software Architect	12	Digital Channels	LeSS, XP	PC3
B3	Project Manager	10	Clinical Systems	DAD, Lean	PC4–PC5
B4	Product Owner	8	Retail Commerce	Scrum, Lean	PC6
B5	Agile Coach	15	Public Services	SAFe, LeSS	PC7

4.5.1. Overall Validation Performance

The tuned hybrid (VIKOR–PROMETHEE, $\alpha = 0.3$, $\beta = 0.7$) achieved MAE = 1.03 and $\rho = 0.53$, a 50.2% reduction in MAE relative to the group-size–weighted per-group single-method baselines. Rank correlations were uniformly positive across framework categories.

4.5.2. Baselines vs. Hybrid

All single-method baselines and hybrid results are computed under the same criteria weight vector for a given comparison. Concretely, for each weighting scheme $\omega \in \{FAHP, Entropy, CRITIC\}$ we compute utilities with PROMETHEE, VIKOR, TOPSIS, EDAS, and M-TOPSIS and then the selected two-engine hybrid. Improvements (Equation (11)) are therefore reported relative to the best single-method baseline under the same ω. Single-method baselines showed mixed convergence: some groups reached moderate positive correlations (e.g., G1 $\rho = 0.62$), while others were negative or inconsistent (e.g., VIKOR in G1 and G3; PROMETHEE and TOPSIS in G5), and MAE often exceeded 2.0 (Table 6). The optimized hybrid improved both MAE and ρ in every group; for G7 ($n = 2$), only MAE is reported. Patterns match practice: Scrum, XP, and Lean lead when adaptability dominates; SAFe, DAD, and APM (in our validation) move up in compliance-driven contexts (cf. Figures 2–4).

Agreement was assessed using a two-way random-effects, absolute-agreement, single-measure intraclass correlation, ICC(2,1). For criteria weighting (Panel A; $k = 5$ raters; $n = 22$ items = O1–O7, P1–P6, C1–C9), ICC(2,1) was 0.76. For framework scoring (Panel B; $k = 5$ raters; $n = 30$ frameworks, each scored on CA, SF, and GAM), ICC(2,1) was 0.68.

To make the tuned setting tangible, Figure 5 shows the compact Hybrid Ranking Lab prototype that updates ranks live as α varies while keeping the expert baseline fixed. Full validation and baseline evidence is reported in the Supplementary Materials (Sections S5 and S6; Tables S12 and S13), including the agreement ledger and the baseline-versus-hybrid comparisons underlying the reported MAE and Spearman ρ.

Table 6. Stage 5: Baseline validation by group, Spearman correlation and MAE (n = 30 frameworks).

		Correlation (Spearman ρ (▲))				Rank Error (MAE (▼))			
Group	**Count**	**PROMETHEE**	**TOPSIS**	**VIKOR**	**Average**	**PROMETHEE**	**TOPSIS**	**VIKOR**	**Average**
G1	7	**0.62**	**0.62**	0.57 ▼	0.22	3.29	3.20	3.29	3.26
G2	3	0.17 ▼	0.06 ▼	0.41	0.06	**0.86**	1.33	1.00	1.06
G3	5	0.59	0.61	0.65 ▼	0.18	2.40	1.20	2.00	1.87
G4	6	0.31	0.44	0.59 ▼	0.05	3.20	2.60	3.00	2.93
G5	4	0.84 ▼	0.82 ▼	**0.84**	0.27 ▼	1.25	1.25	1.25	1.25
G6	3	0.63	0.29	0.27 ▼	0.22	1.30	0.67	1.33	1.10
G7	2	–	–	–	–	1.00	**0.00**	**1.00**	0.67
Weighted Avg	30	0.26	0.27	0.25 ▼	0.09	2.26	1.83	2.17	2.09

Notation. **Bold** = best; red $\rho < 0$ (inverse); ▲ better for ρ; ▼ better for MAE; – = N/A.

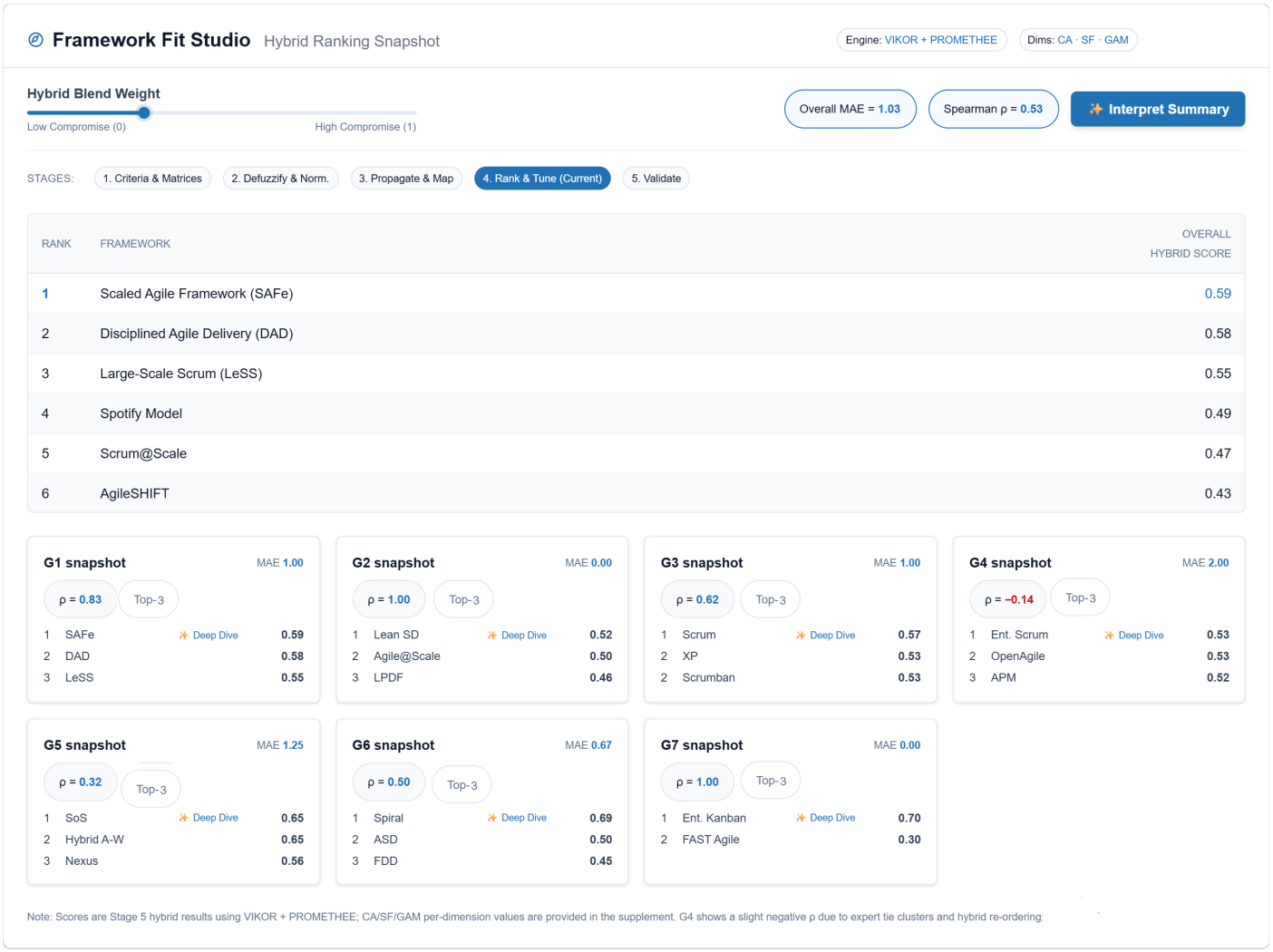


Figure 5. Hybrid ranking interface showing the α–β control ($\alpha + \beta = 1$), CA, SF, and GAM contributions, and expert comparison (MAE, ρ). The baseline at $\alpha = 0.3$, $\beta = 0.7$ reflects the tuned configuration reported in Stage 5.

5. Discussion

Industrial-scale decisions require sufficient governance for viability and compliance, sufficient scalability for portfolio value, and sufficient adaptability for teams. The proposed fuzzy MCDM framework models these trade-offs, optimizes rankings under uncertainty, and provides simple controls and monitoring to keep decisions auditable over time. Separating weighting from ranking, and exposing α–β as a single control lever,

improves transparency without adding operational burden. This section interprets the Stage 3–5 ranking and validation results in managerial and governance terms, highlighting decision trade-offs and actionable selection guidance (with full intermediate artifacts in Supplementary Sections S3–S5).

5.1. Implications for Sustainable Project Management

Rankings show distinct sustainability profiles. Team-centric methods (e.g., Scrum, XP) score strongly on social sustainability (autonomy, well-being) and time-to-value (economic). Governance-oriented families (e.g., SAFe, DAD) can improve enterprise-scale coordination and risk control. Nevertheless, this often entails higher implementation effort (C6) and lower cost efficiency (C9) with constraints on flexibility (C4). The proposed model quantifies trade-offs, enabling organizations to select based on their priorities.

Sustainability is operationalized via the criteria (C1–C9) and their mapping to the super-criteria {CA, SF, GAM}, producing W_{super} that carries economic, social, and environmental considerations (e.g., cost efficiency, autonomy, distributed collaboration). Emphasis among these dimensions is set by the organizational/project pairwise inputs and the 9×3 criteria-to-super mapping (Equation (4)). The α–β setting only calibrates the blend of ranking methods (e.g., VIKOR vs. PROMETHEE) and does not encode domain or sustainability preferences.

5.2. Multi-Level Criteria and Their Influence

Our multi-level criteria separate options that are applicable once the context is explicit. In settings with formal governance and higher maturity, structured frameworks such as SAFe and DAD tend to rank higher [99]. Where requirements are complex and stakeholders are highly engaged, teams often favor lighter methods such as Scrum or XP rather than enterprise-scale frameworks like SAFe or LeSS. At the same time, the Scrum-centric LeSS may still be chosen depending on context [100]. Grouping alternatives by adoption philosophy (Scaling, Lean/Flow, Governance) makes preferences visible, and the criteria reveal the trade-offs across CA, SF, and GAM. Fuzzy prioritization reduces subjectivity compared with purely qualitative assessments [81], thereby satisfying Objective 1. The only negative group correlation (G4, $rho = -0.14$) reflects rank inversions, as experts prioritized strict governance (e.g., RAGE at Exp 1). In contrast, the hybrid balances CA/SF/GAM under $\alpha = 0.3$, $\beta = 0.7$, rewarded options with a stronger agility–governance compromise (for example, Enterprise Scrum to Hybrid 1). Expert ties are expected because five 1–5 ratings are normalized and averaged, so identical means, which could indicate deliberate equivalence rather than a rounding artifact, Scheme 1.

Hypothesis H1 Testing Result

H1: Grouping Agile frameworks helps teams match choices to project and organizational needs.
Outcome: Supported. The group suitability matrix showed stable links between project attributes and framework placement, making the fit easier to explain in context.

Scheme 1. Hypothesis H1 testing result (supported): Agile framework grouping improves contextual fit and explainability.

5.3. Methodological and Theoretical Contributions

We put a simple test on the table: can a hybrid FAHP–MCDM approach outperform any single method against expert rankings? FAHP carries the uncertainty in expert judgments [80]. At the same time, the rankers bring different lenses, TOPSIS for closeness to an

ideal, VIKOR for compromise, PROMETHEE for preference flow, and EDAS for deviation from the mean. The math stays manageable: FAHP weighting scales as $\mathcal{O}(g \times m^3)$ and the hybrid ranking as $\mathcal{O}(n^2)$. Single-method baselines were uneven, some showed negative correlations, and MAE was often above 2.0 (Table 6). The hybrid fixed that issue. VIKOR–PROMETHEE (Entropy) delivered the lowest MAE and the highest ρ (Table 7), which points to reconciling different logics rather than taking a simple average. Comparable hybrid setups have been reported in software-company evaluation and agile-attribute assessments, reinforcing the practical benefits of reconciling utility- and preference-based logics [12,13], Scheme 2.

Hypothesis H2 Testing Result

H2: The hybrid MCDM approach aligns closely with expert rankings.
Outcome: Supported. With VIKOR–PROMETHEE (Entropy), MAE fell to 1.03 and ρ reached 0.53, a 50.2% reduction versus the group-size–weighted per-group single-method baselines.

Scheme 2. Summary of the study's methodological and theoretical contributions of the hybrid FAHP–MCDM approach.

Table 7. Stage 5: Expert versus hybrid rankings (VIKOR–PROMETHEE, $\alpha = 0.3$, $\beta = 0.7$). We report group MAE, the change in MAE (ΔMAE) by method (P, T, V), and Spearman's ρ. For ΔMAE, a negative is an improvement. The "Avg.%" is the mean percent reduction across methods (ignoring any baseline with zero denominator).

Group	F-ID	Framework	Exp Avg	Exp Rank (▼)	Hybrid Score	Hybrid Rank (▼)	Avg. MAE (▼)	Spearman ρ (▲)	ΔMAE (P, T, V)
G1	F2	Disciplined Agile Delivery (DAD)	0.80	3	0.58	2			
G1	F1	Agile Fluency Model	0.72	6	0.39	7			
G1	F4	Large-Scale Scrum (LeSS)	0.81	2	0.55	3			
G1	F5	Scaled Agile Framework (SAFe)	0.83	1	0.59	1	1.00	0.83	(2.29 ▼, 1 ▼, 2.29 ▼) Avg.%: 63.0%
G1	F3	AgileSHIFT	0.79	4	0.43	6			
G1	F6	Scrum@Scale	0.68	7	0.47	5			
G1	F7	Spotify Model	0.79	4	0.49	4			
G2	F8	Lean Software Development	0.80	1	0.52	1			
G2	F9	Lean Product Development Flow	0.68	3	0.46	3	0.00	1.00	(0.67 ▼, 1.33 ▼, 0.67 ▼) Avg.%: 100.0%
G2	F10	Agile@Scale	0.76	2	0.50	2			
G3	F12	Scrum	0.81	1	0.57	1			
G3	F13	XP (Extreme Programming)	0.76	2	0.53	2			
G3	F15	Crystal	0.67	3	0.37	4	1.00	0.62	(1 ▼, 0.00 ○, 1 ▼) Avg.%: 50.0%
G3	F16	Path to Agility	0.63	5	0.41	3			
G3	F28	ScrumBan	0.64	4	0.53	2			
G4	F23	Agile Portfolio Management (APM)	0.69	2	0.52	3			
G4	F26	Enterprise Scrum	0.62	5	0.53	1			
G4	F21	RAGE	0.74	1	0.46	5	2.00	-0.14	(1 ▼, 0.67 ▲, 1 ▼) Avg.%: 11.1%
G4	F24	DA FLEX	0.65	4	0.44	6			
G4	F11	Dynamic Systems Development (DSDM)	0.62	5	0.49	4			
G4	F29	OpenAgile	0.66	3	0.53	2			
G5	F19	Scrum of Scrums (SoS)	0.71	2	0.65	1			
G5	F17	Hybrid Agile-Waterfall	0.68	3	0.65	2	1.25	0.32	(0.00 ○, 0.00 ○, 0.00 ○) Avg.%: 0.0%
G5	F18	Nexus	0.73	1	0.56	3			
G5	F27	Team of Teams	0.64	4	0.35	4			
G6	F20	Spiral Model	0.76	2	0.69	1			
G6	F14	Feature-Driven Development (FDD)	0.67	3	0.45	3	0.67	0.50	(0.67 ▼, 0.00 ○, 0.67 ▼) Avg.%: 50.0%
G6	F30	Adaptive Software Development (ASD)	0.81	1	0.50	2			
G7	F22	Enterprise Kanban	0.65	1	0.70	1	0.00	1.00	(0.5 ▼, 1 ▼, 1 ▼) Avg.%: 100.0%
G7	F25	FAST Agile	0.64	2	0.30	2			
		Overall (weighted, $n = 30$)					1.03	0.53	(1.23 ▼, 0.8 ▼, 1.14 ▼)
		Overall % improvement							50.2%
		Method-wise % improvement							P 54.4%, T 43.7%, V 52.5%

5.4. Dynamic Optimization and Robustness

We asked whether a tunable weighting scheme would make the rankings more adaptable. The α–β control balances the ranking logics with $\alpha + \beta = 1$, and we tune it through

sensitivity checks. In our validation, governance-oriented and team-centric families stayed steady; In our case, the top candidates stayed steady across sweeps; mid-ranked options shifted modestly. The best choice, $\alpha = 0.3$ and $\beta = 0.7$ inside VIKOR–PROMETHEE, cut MAE and lifted ρ, which supports H3, Scheme 3.

Hypothesis H3 Testing Result

H3: The α–β weighting scheme improves adaptability and accuracy of framework rankings.
Outcome: Supported. Using $\alpha = 0.3$ and $\beta = 0.7$ in VIKOR–PROMETHEE produced the lowest MAE and held steady under sensitivity checks.

Scheme 3. Hypothesis H3 testing result (supported): α–β tuning improves ranking accuracy and stability under sensitivity checks.

Our workflow gives managers a fast, auditable way to rank governance, scalability, and flexibility. Fuzzy pairwise weights feed a tuned VIKOR–PROMETHEE mix ($\alpha = 0.3$, $\beta = 0.7$), and sensitivity plots confirm the choices stay stable as priorities change.

5.5. Using the Model in Practice

This model is a decision aid, not a guarantee. It works level by level so that people can stay focused: at any moment, you look at one layer—organization, project, group, or framework—and the most extensive comparison set is the nine shared criteria. That cap is deliberate. It keeps the cognitive load low and leaves room for dialogue. Notes sit next to scores so the story travels with the numbers, and results are read in the context of company size, governance maturity, and who is in the room. The aim is an auditable trail of why a choice fits, not a black box.

There are also times to pause or skip. If a framework is mandated, if the decision is small or time-boxed to a short pilot, if you do not have a few informed practitioners, or if the criteria are still unsettled, the model will not add much. However, it is not advisable to use it to end a debate; use it to tell the story. After delivery, you do not need to revisit all twenty-two items. If you want a quick review, look at the three super-criteria or the nine shared criteria and capture what changed. People still make the decision; the model helps them make it transparently.

5.6. Limitations, Validity, and Future Work

There are three caveats to keep in mind. First, we relied on ten senior practitioners, so generalizability is limited, even if the Delphi rounds helped reduce bias. Moreover, elicited weights and trade-offs reflect the organization's governance culture and domain constraints, so transfer across industries requires re-elicitation rather than direct reuse. Second, the α–β tuning needs software support, and adoption will be easier when the tools are simple to use. Third, the hybrid pipeline is more complex than a single method, but it executes rapidly on a modern laptop, which is practical; even so, it should be tested at larger scales to confirm scalability. Fourth, robustness under extreme or conflicting governance inputs across different organizations was not exhaustively stress-tested, because validation was conducted within a single industrial setting; consequently, the optimal (α, β) pair (here, $\alpha = 0.3, \beta = 0.7$) should be treated as context-specific rather than universal. Accordingly, future studies should replicate the sensitivity grid in additional organizations and deliberately simulate conflicting expert signals to quantify cross-context stability. For cross-industry transfer, the criteria set (if needed), the criteria weights, and the α–β sensitivity grid should be re-run with a panel that matches the target domain's governance and delivery practices; the scoring rubric should also be re-checked for domain-specific

interpretations of CA, SF, and GAM. We focused on TFNs to keep elicitation lightweight; centroid-matched trapezoidal fuzzy numbers and interval/type-2 fuzzy sets are natural extensions when experts express wider ambiguity or disagreement. A systematic cross-shape comparison (TFN vs. trapezoidal vs. type-2) is therefore left for future work. Finally, the inter-level propagation is linear and compensatory, so strengths in some criteria can offset weaknesses in others; in governance-heavy settings (e.g., compliance/safety), veto or dominance effects may be required, which we treat as a non-compensatory extension for future work.

The study deliberately used two expert panels from the same organization to preserve ecological validity between strategic weighting and operational ranking. Although this reduces inter-organizational diversity, it ensures internal consistency between the weighted criteria and the project-specific framework evaluations. Panel composition can also influence outcomes, because role mix, seniority, and functional coverage shape perceived trade-offs; future replications should widen representation across governance, delivery, architecture, and product functions, and include cross-sector panels when benchmarking transferability. For external validity, we reported MAE and correlation, and for reproducibility, we used deterministic code, version control, and archived intermediate artifacts, directly answering current calls for transparency in MCDM [101].

The following steps are pragmatic. Try the framework in compliance-intensive settings such as aerospace and defense [102,103], hook it into DevOps and AI-driven governance pipelines [104], and explore machine-learning-based adaptive weighting [105]. Cross-industry benchmarking is essential to quantify how organizational culture and domain constraints shift the weight profiles and the tuned hybrid setting. Using broader datasets and cross-sector benchmarks will test scale and comparability in practice, turning the framework from a promising prototype into a dependable foundation for next-generation decision support in Agile transformation.

6. Conclusions

Effective scaling should balance delivery speed with long-term sustainability across economic, social, and environmental dimensions. We presented a governance-aware fuzzy MCDM framework that embeds triple-bottom-line considerations into a multi-level evaluation, pairing FAHP for uncertainty-aware weighting with a tunable VIKOR–PROMETHEE ranking stage while keeping the two steps auditable and distinct. In practitioner validation, a tuned setting ($\alpha = 0.3,\ \beta = 0.7$) reduced rank error by 50.2% against group-size weighted single-method baselines, with MAE $= 1.03$ and Spearman $\rho = 0.53$. For practice, the procedure is straightforward: set priorities for governance, scalability, and flexibility; derive fuzzy weights; apply the hybrid ranking; and check stability with sensitivity analysis. Because each step is traceable, selections are easier to explain, audit, and adapt as sustainability requirements evolve. The evidence suggests there is no universal best framework; the method makes trade-offs explicit and defensible for the context at hand.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/pr14020232/s1, Table S1. Core equations and notation used throughout the study. Table S2. Stage 1—Organizational criteria pairwise comparison matrix (O1–O7). Table S3. Stage 1—Organizational-to-project influence matrix (O1–O7 mapped to P1–P6). Table S4. Stage 1—Project-to-group suitability matrix (P1–P6 mapped to G1–G7). Table S5. Stage 1—Group-level suitability profiles across nine criteria (C1–C9), expressed as triangular fuzzy numbers (TFNs). Table S6. TFN mapping Acs: C1–C9 to CA, SF, GAM. Table S7. Stage 2—Framework screening matrix across Core Alignment (CA), Scalability & Flexibility (SF), and Governance & Agile Maturity (GAM). Table S8. Mapping evaluation factors (C*, P*) to triple-bottom-line and SDGs. Table S9. Super-criteria weights across schemes (used in Stage 3 ranking). Table S10. Representative utilities

(subset; FAHP weights). Table S11. Sensitivity performance across α–β (VIKOR–PROMETHEE; Entropy weights). Table S12. Validation ledger: panels, consistency, and agreement. Table S13. Data dictionary for processed artifacts used in reproduction. Table S14. Overview of Agile Frameworks (Part 1: Identifier and Name). Table S15. Overview of Agile Frameworks (Part 2: Strengths, Challenges, and Research Evidence). Figure S1. Prototype screens (S1–S4) for the governance-aware FAHP → VIKOR–PROMETHEE workflow. Top row: S1 Criteria & Provenance Builder; S2 Pairwise & Consistency Console. Bottom row: S3 Hybrid Ranking Lab; S4 Sensitivity & Trade-off Explorer.

Author Contributions: Conceptualization, I.A., A.A.O., M.B. and F.A.; Methodology, I.A., A.A.O., M.B. and F.A.; Software, I.A., A.A.O., M.B. and F.A.; Validation, I.A., A.A.O., M.B. and F.A.; Formal analysis, I.A., A.A.O., M.B. and F.A.; Investigation, I.A., A.A.O., M.B. and F.A.; Resources, I.A., A.A.O., M.B. and F.A.; Data curation, I.A., A.A.O., M.B. and F.A.; Writing—original draft, I.A., A.A.O., M.B. and F.A.; Writing—review & editing, I.A., A.A.O., M.B. and F.A.; Visualization, I.A., A.A.O., M.B. and F.A.; Supervision, I.A., A.A.O., M.B. and F.A.; Project administration, I.A., A.A.O., M.B. and F.A.; Funding acquisition, I.A., A.A.O., M.B. and F.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received partial support from Philadelphia University.

Data Availability Statement: All defuzzified weights, decision matrices, and sensitivity outputs are available in the Supplementary Materials.

Acknowledgments: We thank the practitioner panels for their time and careful judgments. A large language model was used only to help draft and cluster candidate criteria text before expert review; human experts provided all numerical inputs and final decisions.

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

Abbreviations

AHP	Analytic Hierarchy Process
CA	Core Alignment
CRITIC	Criteria Importance Through Inter-criteria Correlation
EDAS	Evaluation based on Distance from Average Solution
FAHP	Fuzzy Analytic Hierarchy Process
GAM	Governance and Agile Maturity
ICC	Intraclass Correlation Coefficient
M-TOPSIS	Modified TOPSIS
MAE	Mean Absolute Error
MCDM	Multi-Criteria Decision-Making
PROMETHEE	Preference Ranking Organization Method for Enrichment Evaluation
SF	Scalability and Flexibility
Spearman ρ	Spearman Rank Correlation Coefficient
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
VIKOR	Multi-Criteria Optimization and Compromise Solution

References

1. Dikert, K.; Paasivaara, M.; Lassenius, C. Challenges and success factors for large-scale agile transformations: A systematic literature review. *J. Syst. Softw.* **2016**, *119*, 87–108. [CrossRef]
2. Kalenda, M.; Hyna, P.; Rossi, B. Scaling agile in large organizations: Practices, challenges, and success factors. *J. Softw. Evol. Process* **2018**, *30*, e1954. [CrossRef]
3. Pinciroli, F. Selection of agile project management approaches based on project complexity. *J. Softw. Evol. Process* **2024**, *36*, e2716. [CrossRef]
4. Eigner, C.; Fahrnberger, G. Challenges in Scaling Agile Frameworks and Ways to Address Them with Scaled Agile Framework (SAFe) and Scrum of Scrums (SoS). *CLEI Electron. J.* **2024**, *27*, 9.

5. Turhan, Y.; Buehrle, D.; Herzwurm, G. Developing a Taxonomy for Agile Scaling Frameworks. In Proceedings of the 7th ACM International Workshop of Software-Intensive Business: Software Business in the Era of Generative Artificial Intelligence (IWSiB '24), Lisbon, Portugal, 16 April 2024; pp. 1–8. [CrossRef]
6. Dutta, P.K.; Bhardwaj, A.K.; Mahida, A. Navigating the Complexities of Agile Transformations in Large Organizations. In *Quantum Computing and Supply Chain Management: A New Era of Optimization*; Hassan, A., Bhattacharya, P., Dutta, P.K., Verma, J.P., Kundu, N.K., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 315–330. [CrossRef]
7. Tsilionis, K.; Ishchenko, V.; Wautelet, Y.; Simonofski, A. Scaling Agility in Large Software Development Projects: A Systematic Literature Review. In *Research and Innovation Forum 2023*; Visvizi, A., Troisi, O., Corvello, V., Eds.; Springer: Cham, Switzerland, 2024; pp. 771–784.
8. Aregbesola, G.D.; Asghar, I.; Akbar, S.; Ullah, R. Fuzzy Logic Model for Informed Decision-Making in Risk Assessment During Software Design. *Systems* **2025**, *13*, 825. [CrossRef]
9. Alshabragi, A.M.; Al-Hajj, A.; Zayed, T. Developing a Maturity Rating System for Project Management Offices. *Systems* **2024**, *12*, 367. [CrossRef]
10. Delesposte, J.E.; Rangel, L.A.D.; Meiriño, M.J.; dos Santos Ferreira, C.M.; Lopes, R.J.F.S.B.; Narcizo, R.B. Model for Innovation Project Selection Supported by Multi-Criteria Methods Considering Sustainability Parameters. *Systems* **2025**, *13*, 876. [CrossRef]
11. Anelli, D.; Morano, P.; Acquafredda, T.; Tajani, F. Structuring Multi-Criteria Decision Approaches for Public Procurement: Methods, Standards and Applications. *Systems* **2025**, *13*, 777. [CrossRef]
12. Torbacki, W. A Framework for Assessing Innovations, Business Models and Sustainability for Software Companies Using Hybrid Multiple-Criteria Decision-Making. *Sustainability* **2024**, *16*, 5871. [CrossRef]
13. Seker, S. Evaluation of Agile Attributes for Low-Cost Carriers to Achieve Sustainable Development Using an Integrated MCDM Approach. *Manag. Decis.* **2024**, *63*, 1229–1261. [CrossRef]
14. Mahmoudian Azar Sharabiani, A.; Mousavi, S.M. A Web-Based Decision Support System for Project Evaluation with Sustainable Development Considerations Based on Two Developed Pythagorean Fuzzy Decision Methods. *Sustainability* **2023**, *15*, 16477. [CrossRef]
15. Fagarasan, C.; Cristea, C.; Cristea, M.; Popa, O.; Pisla, A. Integrating Sustainability Metrics into Project and Portfolio Performance Assessment in Agile Software Development: A Data-Driven Scoring Model. *Sustainability* **2023**, *15*, 13139. [CrossRef]
16. Rehman Khan, S.U.; Younus, M.; Iqbal, J.; Basit Ur Rahim, M.A. A Fuzzy AHP-based Quantitative Framework to Prioritize the Crowd-Based Requirements. In Proceedings of the IEEE 24th International Conference on Software Quality, Reliability, and Security Companion (QRS-C), Cambridge, UK, 1–5 July 2024; pp. 680–691. [CrossRef]
17. Shameem, M.; Kumar, R.R.; Nadeem, M.; Khan, A.A. Taxonomical classification of barriers for scaling agile methods in global software development environment using fuzzy analytic hierarchy process. *Appl. Soft Comput.* **2020**, *90*, 106122. [CrossRef]
18. Neve, J.R.; Agarwal, S. An Interdisciplinary Study of Fuzzy AHP Model for Prioritizing Agile Cost Overhead and Infusion of Machine Learning. In Proceedings of the Asia Pacific Conference on Innovation in Technology (APCIT), Mysore, India, 26–27 July 2024; pp. 1–6. [CrossRef]
19. Camara, R.; Marinho, M.; Moura, H. Agile tailoring in distributed large-scale environments using agile frameworks: A Systematic Literature Review. *CLEI Electron. J.* **2024**, *27*, 8. [CrossRef]
20. Sandhya, K.; Mahapatra, H.B.; Goswami, B.; Acharjya, P.P. Selection of SDM: A fuzzy AHP approach. *Int. J. Comput. Appl.* **2016**, *140* . [CrossRef]
21. Bakhtouchi, A.; Rahmouni, R. A Tree Decision Based Approach for Selecting Software Development Methodology. In Proceedings of the 2018 International Conference on Smart Communications in Network Technologies, SaCoNeT 2018, El Oued, Algeria, 27–31 October 2018; pp. 211–216. [CrossRef]
22. Alemi-Ardakani, M.; Milani, A.S.; Yannacopoulos, S.; Shokouhi, G. On the effect of subjective, objective and combinative weighting in multiple criteria decision making: A case study on impact optimization of composites. *Expert Syst. Appl.* **2016**, *46*, 426–438. [CrossRef]
23. Yue, C.; Huang, R.; Towey, D.; Xian, Z.; Wu, G. An entropy-based group decision-making approach for software quality evaluation. *Expert Syst. Appl.* **2024**, *238*, 121979. [CrossRef]
24. Opricovic, S.; Tzeng, G.H. The Compromise Solution by MCDM Methods: A Comparative Analysis of VIKOR and TOPSIS. *Eur. J. Oper. Res.* **2004**, *156*, 445–455. [CrossRef]
25. Brans, J.P.; Vincke, P. Note–A Preference Ranking Organisation Method. *Manag. Sci.* **1985**, *31*, 647–656. [CrossRef]
26. Liu, Y.; Eckert, C.M.; Earl, C. A review of fuzzy AHP methods for decision-making with subjective judgements. *Expert Syst. Appl.* **2020**, *161*, 113738. [CrossRef]
27. Estrada-Esponda, R.D.; López-Benítez, M.; Matturro, G.; Osorio-Gómez, J.C. Selection of software agile practices using Analytic Hierarchy Process. *Heliyon* **2024**, *10*, e22948. [CrossRef]

28. Khalifeh, A.; Farrell, P.; Alrousan, M.; Alwardat, S.; Faisal, M. Incorporating sustainability into software projects: A conceptual framework. *Int. J. Manag. Proj. Bus.* **2020**, *13*, 1339–1361. [CrossRef]
29. Gomes Silva, F.J.; Kirytopoulos, K.; Pinto Ferreira, L.; Sá, J.C.; Santos, G.; Cancela Nogueira, M.C. The three pillars of sustainability and agile project management: How do they influence each other. *Corp. Soc. Responsib. Environ. Manag.* **2022**, *29*, 1495–1512. [CrossRef]
30. Zakrzewska, M.; Piwowar-Sulej, K.; Jarosz, S.; Sagan, A.; Sołtysik, M. The linkage between Agile project management and sustainable development: A theoretical and empirical view. *Sustain. Dev.* **2022**, *30*, 855–869. [CrossRef]
31. Ezell, B.; Lynch, C.J.; Hester, P.T. Methods for weighting decisions to assist modelers and decision analysists: A review of ratio assignment and approximate techniques. *Appl. Sci.* **2021**, *11*, 10397. [CrossRef]
32. Takeda, E.; Cogger, K.O.; Yu, P.L. Estimating criterion weights using eigenvectors: A comparative study. *Eur. J. Oper. Res.* **1987**, *29*, 360–369. [CrossRef]
33. Giannarou, L.; Zervas, E. Using Delphi technique to build consensus in practice. *Int. J. Bus. Sci. Appl. Manag. (IJBSAM)* **2014**, *9*, 65–82. [CrossRef]
34. Hsu, C.C.; Sandford, B.A. The Delphi technique: Making sense of consensus. *Pract. Assess. Res. Eval.* **2007**, *12*, 10.
35. Rezaei, J. Best-worst multi-criteria decision-making method. *Omega* **2015**, *53*, 49–57. [CrossRef]
36. Diakoulaki, D.; Mavrotas, G.; Papayannakis, L. Determining objective weights in multiple criteria problems: The critic method. *Comput. Oper. Res.* **1995**, *22*, 763–770. [CrossRef]
37. Gao, X.; Ma, Y.; Zhou, W. Analysis of Software Trustworthiness Based on FAHP-CRITIC Method. *J. Shanghai Jiaotong Univ. (Science)* **2024**, *29*, 588–600. [CrossRef]
38. Niepostyn, S.J.; Daszczuk, W.B. Entropy as a Measure of Consistency in Software Architecture. *Entropy* **2023**, *25*, 328. [CrossRef]
39. Bao, Q.; Yuxin, Z.; Yuxiao, W.; Feng, Y. Can entropy weight method correctly reflect the distinction of water quality indices? *Water Resour. Manag.* **2020**, *34*, 3667–3674. [CrossRef]
40. Saaty, T.L. A scaling method for priorities in hierarchical structures. *J. Math. Psychol.* **1977**, *15*, 234–281. [CrossRef]
41. Saaty, T.L. How to make a decision: The analytic hierarchy process. *Eur. J. Oper. Res.* **1990**, *48*, 9–26. [CrossRef]
42. Saaty, T.L. Making and validating complex decisions with the AHP/ANP. *J. Syst. Sci. Syst. Eng.* **2005**, *14*, 1–36. [CrossRef]
43. Divya; Anand, A.; Johri, P.; Papic, L. Chapter 10—AHP based determination of critical testing coverage measures for reliable & complex software systems. In *Reliability Assessment and Optimization of Complex Systems*; Kumar, A., Bhandari, A.S., Ram, M., Eds.; Advances in Reliability Science; Elsevier: Amsterdam, The Netherlands, 2025; pp. 191–218. [CrossRef]
44. Nazim, M.; Wali Mohammad, C.; Sadiq, M. A comparison between fuzzy AHP and fuzzy TOPSIS methods to software requirements selection. *Alex. Eng. J.* **2022**, *61*, 10851–10870. [CrossRef]
45. Nadeem, M. Analyze quantum security in software design using fuzzy-AHP. *Int. J. Inf. Technol.* **2024**, *17*, 5563–5575. [CrossRef]
46. Alyamani, R.; Long, S. The Application of Fuzzy Analytic Hierarchy Process in Sustainable Project Selection. *Sustainability* **2020**, *12*, 8314. [CrossRef]
47. Hwang, C.L.; Yoon, K. *Multiple Attribute Decision Making: Methods and Applications*; Lecture Notes in Economics and Mathematical Systems; Springer: Berlin/Heidelberg, Germany, 1981; Volume 186. [CrossRef]
48. Hwang, C.L.; Lai, Y.J.; Liu, T.Y. A new approach for multiple objective decision making. *Comput. Oper. Res.* **1993**, *20*, 889–899. [CrossRef]
49. Yoon, K. A reconciliation among discrete compromise situations. *J. Oper. Res. Soc.* **1987**, *38*, 277–286. [CrossRef]
50. Pandey, V.; Komal; Dincer, H. A review on TOPSIS method and its extensions for different applications with recent development. *Soft Comput.* **2023**, *27*, 18011–18039. [CrossRef]
51. Verma, S.; Mehlawat, M.K.; Mahajan, D. Software component evaluation and selection using TOPSIS and fuzzy interactive approach under multiple applications development. *Ann. Oper. Res.* **2022**, *312*, 441–471. [CrossRef]
52. Akram, F.; Ahmad, T.; Sadiq, M. An integrated fuzzy adjusted cosine similarity and TOPSIS based recommendation system for information system requirements selection. *Decis. Anal. J.* **2024**, *11*, 100443. [CrossRef]
53. Anbarkhan, S.H. A Fuzzy-TOPSIS-Based Approach to Assessing Sustainability in Software Engineering: An Industry 5.0 Perspective. *Sustainability* **2023**, *15*, 13844. [CrossRef]
54. Mateen Khan, F.; Munir, A.; Albaity, M.; Nadeem, M.; Mahmood, T. Software Reliability Growth Model Selection by Using VIKOR Method Based on q-Rung Orthopair Fuzzy Entropy and Divergence Measures. *IEEE Access* **2024**, *12*, 86572–86582. [CrossRef]
55. Meng, X.; Lu, Y.; Liu, J. A risk evaluation model of electric power cloud platform from the information perspective based on fuzzy type-2 VIKOR. *Comput. Ind. Eng.* **2023**, *184*, 109616. [CrossRef]
56. Lai, H.; Liao, H.; Wen, Z.; Zavadskas, E.K.; Al-Barakati, A. An improved CoCoSo method with a maximum variance optimization model for cloud service provider selection. *Eng. Econ.* **2020**, *31*, 411–424. [CrossRef]

57. Keshavarz Ghorabaee, M.; Zavadskas, E.K.; Olfat, L.; Turskis, Z. Multi-Criteria Inventory Classification Using a New Method of Evaluation Based on Distance from Average Solution (EDAS). *Informatica* **2015**, *26*, 435–451. [CrossRef]
58. Torkayesh, A.E.; Deveci, M.; Karagoz, S.; Antucheviciene, J. A state-of-the-art survey of evaluation based on distance from average solution (EDAS): Developments and applications. *Expert Syst. Appl.* **2023**, *221*, 119724. [CrossRef]
59. Rathor, S.; Agrawal, S.C. Wireless Network Environment Evaluation using EDAS Method. In Proceedings of the 1st International Conference on Cognitive Computing and Engineering Education (ICCCEE), Pune, India, 27–29 April 2023; pp. 1–5. [CrossRef]
60. Behzadian, M.; Kazemzadeh, R.B.; Albadvi, A.; Aghdasi, M. PROMETHEE: A comprehensive literature review on methodologies and applications. *Eur. J. Oper. Res.* **2010**, *200*, 198–215. [CrossRef]
61. Brans, J.P. *L'ingénierie de la Décision: Élaboration D'instruments D'aide à la Décision. La Méthode PROMETHEE*; Presses de l'Université Laval: Quebec, QC, Canada, 1982. (In French)
62. Caloğlu, Z.V.; Zontul, M.; Yemen, I.; Bağrıyanik, S. Software Quality Measurement Modelling Using AHP and PROMETHEE Methods. In Proceedings of the 6th International Conference on Computer Science and Engineering (UBMK), Ankara, Turkey, 15–17 September 2021; pp. 608–612. [CrossRef]
63. Aruchsamy, R.; Velusamy, I.; Sanmugavel, K.; Dhandapani, P.B.; Ramasamy, K. Generalization of Fermatean Fuzzy Set and Implementation of Fermatean Fuzzy PROMETHEE II Method for Decision Making via PROMETHEE GAIA. *Axioms* **2024**, *13*, 408. [CrossRef]
64. Chisale, S.W.; Eliya, S.; Taulo, J. Optimization and design of hybrid power system using HOMER pro and integrated CRITIC-PROMETHEE II approaches. *Green Technol. Sustain.* **2023**, *1*, 100005. [CrossRef]
65. Yazdani, M.; Zarate, P.; Zavadskas, E.K.; Turskis, Z. A Combined Compromise Solution (CoCoSo) Method for Multi-Criteria Decision-Making Problems. *Manag. Decis.* **2019**, *57*, 2501–2519. [CrossRef]
66. Yu, D.; Liu, Y.; Xu, Z. Analysis of knowledge evolution in PROMETHEE: A longitudinal and dynamic perspective. *Inf. Sci.* **2023**, *642*, 119151. [CrossRef]
67. Roselli, L.R.P.; de Almeida, A.T. FITradeoff Decision Support System Applied to Solve a Supplier Selection Problem. In *Advances in Information Systems, Artificial Intelligence and Knowledge Management*; Saad, I., Rosenthal-Sabroux, C., Gargouri, F., Chakhar, S., Williams, N., Haig, E., Eds.; Springer: Cham, Switzerland, 2024; pp. 49–62.
68. Hezam, I.M.; Rani, P.; Mishra, A.R.; Alshamrani, A. Assessment of autonomous smart wheelchairs for disabled persons using hybrid interval-valued Fermatean fuzzy combined compromise solution method. *Sustain. Energy Technol. Assess.* **2023**, *57*, 103169. [CrossRef]
69. Figueira, J.; Greco, S.; Ehrgott, M. (Eds.) *Multiple Criteria Decision Analysis: State of the Art Surveys*; International Series in Operations Research & Management Science; Springer: New York, NY, USA, 2005; Volume 78. [CrossRef]
70. Leong, W.P.; Tan, C.W.; Singh, G. A New Integrated Multi-Criteria Decision-Making Model for Resilient Supplier Selection. *Appl. Syst. Innov.* **2022**, *5*, 8. [CrossRef]
71. Shih, H.S.; Shyur, H.J.; Lee, E.S. An extension of TOPSIS for group decision making. *Math. Comput. Model.* **2007**, *45*, 801–813. [CrossRef]
72. Vaidya, O.S.; Kumar, S. Analytic hierarchy process: An overview of applications. *Eur. J. Oper. Res.* **2006**, *169*, 1–29. [CrossRef]
73. Basilio, M.P.; Pereira, V.; Yiğit, F. New hybrid EC-PROMETHEE method with multiple iterations of random weight ranges: Step-by-step application in Python. *MethodsX* **2024**, *13*, 102890. [CrossRef] [PubMed]
74. Khan, S.; Purohit, L. An Integrated Methodology of Ranking Based on PROMETHEE-CRITIC and TOPSIS-CRITIC In Web Service Domain. In Proceedings of the IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Indore, India, 23–24 April 2022; pp. 335–340. [CrossRef]
75. Mutambik, I. The Sustainability of Smart Cities: Improving Evaluation by Combining MCDA and PROMETHEE. *Land* **2024**, *13*, 1471. [CrossRef]
76. Shemshadi, A.; Shirazi, H.; Toreihi, M.; Tarokh, M.J. A fuzzy VIKOR method for supplier selection based on entropy measure for objective weighting. *Expert Syst. Appl.* **2011**, *38*, 12160–12167. [CrossRef]
77. Sheykhizadeh, M.; Ghasemi, R.; Vandchali, H.R.; Sepehri, A.; Torabi, S.A. A Hybrid Decision-Making Framework for a Supplier Selection Problem Based on Lean, Agile, Resilience, and Green Criteria: A Case Study of a Pharmaceutical Industry. *Environ. Dev. Sustain.* **2024**, *26*, 30969–30996. [CrossRef]
78. Bhol, S.G. Applications of Multi Criteria Decision Making Methods in Cyber Security. In *Cyber-Physical Systems Security: A Multi-disciplinary Approach*; Choudhury, A., Kaushik, K., Kumar, V., Singh, B.K., Eds.; Springer Nature: Singapore, 2025; pp. 233–258. [CrossRef]
79. Yue, C.; Huang, R.; Towey, D.; Zhou, L. A median-based fuzzy approach to software quality evaluation. *Tsinghua Sci. Technol.* **2024**, *30*, 2146–2168. [CrossRef]

80. De Wilde, L.; Macharis, C.; Keseru, I. Technical requirements for organising successful mobility campaigns in citizen observatories. *Transp. Res. Procedia* **2020**, *48*, 1418–1429. [CrossRef]
81. Alqudah, M.K.; Razali, R.; Alqudah, M.K. Agile Methods Selection Model: A Grounded Theory Study. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 357. [CrossRef]
82. Beecham, S.; Clear, T.; Lal, R.; Noll, J. Do scaling agile frameworks address global software development risks? An empirical study. *J. Syst. Softw.* **2021**, *171*, 110823. [CrossRef]
83. Verwijs, C.; Russo, D. Do Agile scaling approaches make a difference? An empirical comparison of team effectiveness across popular scaling approaches. *Empir. Softw. Eng.* **2024**, *29*, 75. [CrossRef]
84. Donaldson, L. *The Contingency Theory of Organizations*; Foundations for Organizational Science; SAGE Publications: Thousand Oaks, CA, USA, 2001. [CrossRef]
85. Zhou, L.; Asano, K. A Two-Layer Model for Complex Multi-Criteria Decision-Making and Its Application in Institutional Research. *Appl. Syst. Innov.* **2025**, *8*, 148. [CrossRef]
86. Katushabe, E.; Mugisha, G.; Ssenyonga, J. Fuzzy-Based Prediction Model for Air Quality Monitoring for Kampala City. *Appl. Syst. Innov.* **2021**, *4*, 44. [CrossRef]
87. Nguyen-Duc, A.; Cabrero-Daniel, B.; Przybylek, A.; Arora, C.; Khanna, D.; Herda, T.; Rafiq, U.; Melegati, J.; Guerra, E.; Kemell, K.K.; et al. Generative Artificial Intelligence for Software Engineering—A Research Agenda. *Softw. Pract. Exp.* **2025**, *55*, 1806–1843. [CrossRef]
88. Lee, K.H.; Choi, G.H.; Yun, J.; Choi, J.; Goh, M.J.; Sinn, D.H.; Jin, Y.J.; Kim, M.A.; Yu, S.J.; Jang, S.; et al. Machine learning-based clinical decision support system for treatment recommendation and overall survival prediction of hepatocellular carcinoma: A multi-center study. *npj Digit. Med.* **2024**, *7*, 2. [CrossRef]
89. Almalki, S.S. AI-Driven Decision Support Systems in Agile Software Project Management: Enhancing Risk Mitigation and Resource Allocation. *Systems* **2025**, *13*, 208. [CrossRef]
90. Ebert, C.; Paasivaara, M. Scaling agile. *IEEE Softw.* **2017**, *34*, 98–103. [CrossRef]
91. Conboy, K.; Carroll, N. Implementing large-scale agile frameworks: Challenges and recommendations. *IEEE Softw.* **2019**, *36*, 44–50. [CrossRef]
92. Demir, G.; Chatterjee, P.; Pamucar, D. Sensitivity analysis in multi-criteria decision making: A state-of-the-art research perspective using bibliometric analysis. *Expert Syst. Appl.* **2024**, *237*, 121660. [CrossRef]
93. Mondal, S.; Bappon, S.D.; Roy, C.K. Enhancing User Interaction in ChatGPT: Characterizing and Consolidating Multiple Prompts for Issue Resolution. In Proceedings of the IEEE/ACM 21st International Conference on Mining Software Repositories (MSR), Lisbon, Portugal, 15–16 April 2024; pp. 222–226.
94. Hou, X.; Zhao, Y.; Liu, Y.; Yang, Z.; Wang, K.; Li, L.; Luo, X.; Lo, D.; Grundy, J.; Wang, H. Large Language Models for Software Engineering: A Systematic Literature Review. *ACM Trans. Softw. Eng. Methodol.* **2024**, *33*, 220:1–220:79. [CrossRef]
95. Alenezi, M.; Akour, M. AI-Driven Innovations in Software Engineering: A Review of Current Practices and Future Directions. *Appl. Sci.* **2025**, *15*, 1344. [CrossRef]
96. Sainio, K.; Abrahamsson, P.; Ahtee, T. Prompt Patterns for Agile Software Project Managers: First Results. In Proceedings of the Software Business, Lahti, Finland, 27–29 November 2023; Hyrynsalmi, S., Münch, J., Smolander, K., Melegati, J., Eds.; Springer: Cham, Switzerland, 2024; pp. 190–204.
97. Nguyen-Duc, A.; Khanna, D. Value-Based Adoption of ChatGPT in Agile Software Development: A Survey Study of Nordic Software Experts. In *Generative AI for Effective Software Development*; Nguyen-Duc, A., Abrahamsson, P., Khomh, F., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 257–273. [CrossRef]
98. Schmidt, D.C.; Spencer-Smith, J.; Fu, Q.; White, J. Towards a Catalog of Prompt Patterns to Enhance the Discipline of Prompt Engineering. *Ada Lett.* **2024**, *43*, 43–51. [CrossRef]
99. Nägele, S.; Schenk, N.; Matthes, F. The Current State of Security Governance and Compliance in Large-Scale Agile Development: A Systematic Literature Review and Interview Study. In Proceedings of the IEEE 25th Conference on Business Informatics (CBI), Prague, Czechia, 21–23 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–10.
100. McHugh, M.; McCaffery, F.; Casey, V. Barriers to adopting agile practices when developing medical device software. In Proceedings of the Software Process Improvement and Capability Determination: 12th International Conference, SPICE 2012, Palma, Spain, 29–31 May 2012; Proceedings 12; Springer: Berlin/Heidelberg, Germany, 2012; pp. 141–147.
101. Madanchian, M.; Taherdoost, H. Applications of Multi-Criteria Decision Making in Information Systems for Strategic and Operational Decisions. *Computers* **2025**, *14*, 208. [CrossRef]
102. Kelecş, N. A comparative evaluation of multi-criteria decision-making framework for armed unmanned aerial vehicle. *Int. J. Intell. Unmanned Syst.* **2024**, *12*, 433–453. [CrossRef]
103. Nain, A.; Jain, D.; Trivedi, A. Multi-criteria decision-making methods: Application in humanitarian operations. *Benchmarking Int. J.* **2024**, *31*, 2090–2128. [CrossRef]

104. Huang, H.; Burgherr, P. MCDA Calculator: A Streamlined Decision Support System for Multi-criteria Decision Analysis. In Proceedings of the International Conference on Decision Support System Technology, Porto, Portugal, 3–5 June 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 31–45. [CrossRef]
105. Abdulla, A.; Baryannis, G. A hybrid multi-criteria decision-making and machine learning approach for explainable supplier selection. *Supply Chain Anal.* **2024**, *7*, 100074. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.