

Keywords: ELECTRE III, IVHFLTS, fuzzy decision-making, AI logistics, threshold sensitivity analysis

Fadoua TAMTAM ^{1*}, Amina TOURABI ¹

¹ National School of Applied Sciences, Systems Engineering and Decision Support Laboratory (LISAD), IBN ZOHR University, Morocco, fadoua.tamtam@gmail.com, a.tourabi@uiz.ac.ma

* Corresponding author: fadoua.tamtam@gmail.com

Enhanced ELECTRE III method with interval-valued hesitant fuzzy linguistic sets for multi-criteria group decision-making in smart supply networks

Abstract

This study presents a robust decision-making framework for evaluating strategic artificial intelligence (AI) initiatives within DHL's smart supply network. The objective is to prioritize four AI alternatives—autonomous warehouse routing, predictive delivery optimization, AI-driven demand forecasting, and intelligent inventory rebalancing—based on eight strategic criteria, including cybersecurity, adaptability, and infrastructure readiness. A cross-functional panel of experts provided linguistic assessments, modeled using Interval-Valued Hesitant Fuzzy Linguistic Term Sets (IVHFLTS) to capture hesitation and uncertainty. These inputs were aggregated and processed by an extended ELECTRE III method incorporating fuzzy thresholds for indifference, preference, and veto. Sensitivity analysis confirmed the stability of the final ranking under $\pm 10\%$ threshold variation, while consensus evaluation revealed expert divergence, which was mitigated by dynamic reweighting. Predictive delivery optimization and intelligent inventory rebalancing emerged as the top-ranked initiatives, aligning with DHL's strategic goals of customer responsiveness and operational resilience. The methodology demonstrates high robustness, interpretability, and practical relevance for AI-driven logistics planning.

1. INTRODUCTION

The logistics and supply chain industry is undergoing a profound transformation, driven by the rapid integration of artificial intelligence (AI) technologies (Shamsuddoha et al., 2025). From predictive analytics and autonomous routing to intelligent inventory systems and real-time decision support, AI is redefining how global logistics companies manage complexity, uncertainty, and customer expectations (Iseri et al., 2025). For multinational logistics providers like DHL, the strategic use of AI is no longer an option - it is a competitive imperative. However, the selection and prioritization of AI initiatives remains a formidable challenge due to the multidimensional nature of these technologies, the diversity of stakeholder perspectives, and the inherent uncertainty surrounding their implementation and impact.

Strategic decision-making in this context requires a robust evaluation framework that can incorporate both quantitative performance indicators and qualitative expert judgments. Traditional cost-benefit analysis or purely financial metrics often fall short in capturing the broader implications of AI deployment, such as such as cybersecurity risks, workforce adaptability, regulatory compliance, and integration with legacy systems (Åström et al., 2022; Pagliaro, 2025; Vuković et al., 2025). In addition, the subjective and often hesitant nature of expert judgments - especially when dealing with emerging technologies - necessitates a modeling approach that can handle linguistic ambiguity and hesitation.

To address these challenges, this study proposes an improved decision framework that integrates the ELECTRE III outranking method with Interval-Valued Hesitant Fuzzy Linguistic Term Sets (IVHFLTS). ELECTRE III is particularly well-suited for complex, no compensatory decision environments where trade-offs must be explicitly modeled through indifference, preference, and veto thresholds (Battisti, 2022; Parolin et al., 2025). Incorporating IVHFLTS allows experts to express their assessments using multiple linguistic terms with associated confidence levels, preserving the richness of subjective input while allowing for rigorous aggregation and comparison (Carnia et al., 2025; Malik et al., 2018).

The study focuses on evaluating four AI initiatives within DHL's Smart Supply Network: autonomous warehouse routing, predictive delivery optimization, AI-driven demand forecasting, and smart inventory rebalancing. These alternatives are evaluated against eight strategic criteria, including cybersecurity robustness, infrastructure readiness, explainability, and customer experience improvement. A panel of experts in logistics operations, AI systems, innovation management, and sustainability provided linguistic evaluations that were aggregated using the IVHFLTS-ELECTRE III framework.

The objectives of this research are threefold: (1) to demonstrate the applicability of ELECTRE III-IVHFLTS in evaluating strategic AI initiatives under uncertainty; (2) to assess the robustness of the final ranking through sensitivity analysis and expert consensus modeling; and (3) to provide actionable insights for strategic planning and investment prioritization in AI-driven logistics transformation.

By combining methodological rigor with practical relevance, this study contributes to the growing body of research on fuzzy multicriteria decision making and provides a replicable framework for evaluating AI initiatives in complex organizational settings.

2. LITERATURE REVIEW

The evaluation of AI initiatives in logistics intersects several research areas, including strategic technology management, multi-criteria decision analysis (MCDA), and fuzzy logic modeling. This section reviews the state of the art in each of these areas, with a particular focus on the integration of ELECTRE III and IVHFLTS in decision making under uncertainty.

2.1. Strategic AI Deployment in logistics

AI technologies are increasingly embedded in logistics operations, enabling companies to optimize routing, forecast demand, automate warehousing, and improve customer experience (Chen et al., 2024; Riad et al., 2024). These applications are not only operational tools, but also strategic assets that influence competitiveness, agility, and resilience. However, the strategic use of AI requires careful evaluation of both tangible and intangible factors, including implementation risk, regulatory alignment, and organizational readiness.

Several studies have proposed frameworks for assessing AI readiness and prioritization in logistics. Montella et al. (2025) emphasized the need for multidimensional evaluation models that integrate technical, economic, and organizational criteria. Their work highlighted the limitations of single-criteria approaches and called for more holistic decision frameworks. Zare et al. (2025) developed a fuzzy AHP-TOPSIS model to assess AI readiness in warehouse automation, demonstrating that hybrid models can effectively capture expert preferences and strategic trade-offs. Aydin et al. (2025) applied a multi-layered fuzzy approach to the evaluation of AI-driven logistics platforms, showing that expert disagreement and linguistic ambiguity are common in strategic evaluations.

Despite these contributions, most existing models rely on additive or compensatory logic, which may not reflect the real-world constraints and trade-offs involved in AI deployment. This limitation underscores the need for non-compensatory methods such as ELECTRE III, which can model situations where poor performance on one criterion cannot be offset by excellence on another.

2.2. Multi-criteria decision analysis in uncertain environments

MCDA provides a structured approach to evaluating alternatives across multiple, often conflicting, criteria. Among the various MCDA methods, outranking approaches such as ELECTRE III are particularly well suited for strategic decision making in complex environments.

ELECTRE III has been widely applied in areas such as transportation planning, infrastructure investment, and environmental management. For example, Yusoff et al. (2023) applied a fuzzy ELECTRE III model to the evaluation of intelligent transportation systems, demonstrating how its non-compensatory logic effectively captures complex trade-offs among conflicting criteria. However, conventional implementations of ELECTRE III often rely on crisp numerical inputs, which may inadequately reflect the hesitation and ambiguity present in expert judgments—especially in the context of emerging technologies such as AI.

To address this limitation, researchers have developed fuzzy extensions to ELECTRE III. These include models based on fuzzy sets, intuitionistic fuzzy sets, and hesitant fuzzy linguistic term sets, which allow for a

more expressive and realistic representation of expert assessments (Hashemi et al., 2016; Veeramachaneni & Kandikonda, 2016). Shao et al. (2023) applied a fuzzy ELECTRE III-IVHFLTS framework to evaluate strategic initiatives in smart city development, demonstrating that integrating fuzzy logic into MCDA enhances decision transparency, stakeholder alignment, and interpretability.

2.3. Interval-valued hesitant fuzzy linguistic term sets (IVHFLTS)

IVHFLTS represent a significant advance in fuzzy decision modeling, allowing experts to express multiple linguistic terms with associated confidence levels and degrees of hesitation. This approach preserves the richness of linguistic input while allowing rigorous aggregation and comparison, making it particularly suitable for group decision making in uncertain environments (Faizi et al., 2024; Liu & Peng, 2023).

Khan et al. (2024) introduced fundamental distance and similarity measures for hesitant fuzzy sets, facilitating their integration into MCDA frameworks. Ruan et al. (2024) extended ELECTRE III with probabilistic interval-valued intuitionistic hesitant fuzzy information to evaluate supply chain resiliency strategies. Their model captured both individual hesitation and group consensus, resulting in more stable and interpretable rankings. Zeng et al. (2023) applied IVHFLTS to smart city planning and demonstrated that the method improved transparency and stakeholder engagement.

Despite these methodological advances, applications of ELECTRE III-IVHFLTS in real-world logistics organizations remain limited. Most studies focus on theoretical development or simulated environments. This study addresses this gap by applying the model to DHL's strategic AI initiatives and validating its robustness through sensitivity testing and expert consensus analysis.

3. METHODS

3.1. Study design

This study was conducted within the strategic innovation department of DHL's global logistics operations, with a focus on evaluating next-generation AI initiatives in intelligent supply networks. The term "initiative" refers to a structured set of AI-enabled actions aimed at improving supply chain responsiveness, resilience, and sustainability. For consistency with the Multi-Criteria Decision Analysis (MCDA) framework, the term "alternative" is used throughout this paper.

Four cross-functional planning teams-Logistics Optimization, AI Systems, Innovation Strategy, and Sustainability Integration-participated in the evaluation process. Each team consisted of three to six members, including operational managers, data scientists, and strategic planners. Two executive-level decision makers initiated the study to assess the strategic viability of their current portfolio of AI initiatives, posing the following decision questions: which alternatives should be prioritized for full deployment; which should be discontinued or postponed; which are currently active but may require strategic repositioning; and which are stable and do not require immediate action.

The evaluation process was designed to support these decision makers by generating structured decision insights through team-led assessments. Each team conducted its own evaluation independently, and no cross-team synthesis or ranking was imposed. The range of alternatives included autonomous warehouse routing, predictive delivery optimization, AI-driven demand forecasting, and intelligent inventory rebalancing. The technical details of each initiative are not detailed here, as the focus of this study is on the methodological framework rather than the operational specifics of each alternative.

A total of 18 participants, including two senior decision makers and 16 team members, contributed to the study. All participants provided informed consent in accordance with DHL's internal research ethics protocol. The study spanned three months and included several virtual workshops, asynchronous rating sessions, and structured elicitation sessions. The lead analyst facilitated the process, guiding the teams through the evaluation framework and implementing the enhanced ELECTRE III methodology with Interval-Valued Hesitant Fuzzy Linguistic Term Sets (IVHFLTS). Each team was responsible for evaluating the alternatives and assigning weights to the evaluation criteria, with methodological support provided through structured templates and three 90-minute virtual sessions per team. The final recommendations were consolidated in a decision workshop led by the executive sponsors.

3.2. Setting objectives and criteria

The evaluation framework was anchored in DHL's strategic goals for the use of AI in logistics. After a series of scoping meetings, decision makers agreed on four overarching objectives: strategic alignment, operational feasibility, technological robustness, and stakeholder impact. These objectives were operationalized through eight evaluation criteria, each selected to represent a distinct dimension of value and risk in the deployment of the AI initiative.

The criteria were defined as follows

- g_1 : Implementation Risk
- g_2 : Data Infrastructure Readiness
- g_3 : Workforce Adaptability
- g_4 : Cybersecurity Robustness
- g_5 : Customer Experience Enhancement
- g_6 : AI Explainability & Transparency
- g_7 : Regulatory Compliance Potential
- g_8 : Integration with Legacy Systems

Each criterion was associated with a qualitative measure expressed by linguistic terms mapped to interval-valued fuzzy sets. The criteria were reviewed to avoid redundancy and ensure conceptual independence. The final set was validated by all participating teams and documented in the evaluation protocol.

3.3. Rating alternatives

Each planning team was asked to rate the expected impact of the four AI initiatives on the eight criteria using a linguistic scale: Very Low, Low, Moderate, High, and Very High. These ratings were expressed as IVHFLTS, allowing team members to capture uncertainty and hesitation in their assessments. For example, a team might rate the robustness of an initiative's cybersecurity as a combination of "high" and "very high," with associated membership levels reflecting confidence levels.

Teams were also asked to document their assumptions and rationale for each rating. These deliberations took place in facilitated sessions and asynchronous exchanges, resulting in a consensus set of IVHFLTS ratings for each alternative. The use of fuzzy linguistic modeling was essential given the complexity of the initiatives and the qualitative nature of many of the criteria. Quantitative scoring was deemed impractical due to the emergent nature of the technologies and the limited availability of performance data.

The aggregated scores were compiled into a fuzzy impact matrix that served as the input for the ELECTRE III outranking analysis. This matrix preserved the richness of expert judgment while allowing for a structured comparison of alternatives.

3.4. Weighting criteria

Criteria weighting was conducted using a structured elicitation protocol based on the SMART method. Each team was asked to assign relative importance to the eight criteria, interpreting the weights as scaling constants reflecting the contribution of each criterion to the overall desirability of an alternative. Weights were normalized to ensure comparability across teams and consistency with the ELECTRE III aggregation model.

The elicitation process included pairwise comparisons, ranking exercises, and facilitated discussions. Final weights were calculated as an average of team-level inputs, with adjustments made to reflect executive priorities. For example, cybersecurity robustness and regulatory compliance received higher weights due to recent strategic directives, while workforce adaptability received a lower weight for teams with mature automation capabilities.

The resulting weight vector $W = \{w_1, w_2, \dots, w_8\}$ was used to calculate concordance indices and credibility scores in the ELECTRE III model. These weights ensured that the assessment reflected DHL's strategic priorities and operational realities.

Tab. 1. Objectives, Criteria, and Measures Used in the Case Study

Objective	Criterion	Measure Description	Best Value	Worst Value
Strategic Alignment	g_1 : Implementation Risk	Degree of uncertainty and operational disruption expected during rollout	Very Low Risk	Very High Risk
	g_2 : Data Infrastructure Readiness	Compatibility with existing data architecture and cloud systems	Fully Compatible	Not Compatible
Operational Feasibility	g_3 : Workforce Adaptability	Ease of adoption and training across logistics teams	Seamless Integration	Major Resistance
	g_4 : Cybersecurity Robustness	Strength of security protocols and resilience against cyber threats	Military-Grade Secure	Vulnerable
Stakeholder Impact	g_5 : Customer Experience Enhancement	Improvement in delivery accuracy, personalization, and service responsiveness	Transformational Impact	No Impact
	g_6 : AI Explainability & Transparency	Clarity of AI decision logic and interpretability for stakeholders	Fully Transparent	Opaque/Black Box
Regulatory & Technical Compliance	g_7 : Regulatory Compliance Potential	Alignment with current and anticipated logistics regulations	Fully Compliant	Non-Compliant
	g_8 : Integration with Legacy Systems	Ability to interface with existing ERP, WMS, and TMS platforms	Plug-and-Play	Requires Full Overhaul

4. RESULTS AND DISCUSSION

4.1. Results

Each expert $e_k \in E$ provided evaluations $x_{ijk} = \tilde{L}_{ijk}$ for each AI initiative $a_i \in A$ under each criterion $g_j \in \mathcal{G}$, expressed as an IVHFLTS (Naqvi et al., 2023):

$$\tilde{L}_{ijk} = \{([l_{ijk}^{(1)}, u_{ijk}^{(1)}], \mu_{ijk}^{(1)}), \dots, ([l_{ijk}^{(t)}, u_{ijk}^{(t)}], \mu_{ijk}^{(t)})\} \quad (1)$$

where: $[l_{ijk}^{(t)}, u_{ijk}^{(t)}] \subseteq [0,1]$ – the interval-valued representation of the linguistic term selected by expert e_k ,
 $\mu_{ijk}^{(t)} \in [0,1]$ – the membership degree (hesitation weight),
 t – the number of hesitant terms provided.

To ensure semantic consistency and methodological validity, the language scale was adopted from the established IVHFLTS literature (Liao et al., 2023; Lima-Junior et al., 2023). The five-point scale is represented as follows:

Tab. 2. The linguistic scale

Linguistic Term	Interval Mapping
Very Low	[0.00, 0.18]
Low	[0.19, 0.38]
Moderate	[0.39, 0.62]
High	[0.63, 0.82]
Very High	[0.83, 1.00]

This mapping preserves ordinal semantics, avoids overlap, and supports centroid-based aggregation. Each expert's input was aggregated using a fuzzy weighted averaging operator:

$$\tilde{L}_{ij} = \bigoplus_{k=1}^4 w_k \otimes \tilde{L}_{ijk}, \text{ with } w_k = 0.25 \quad (2)$$

The scalar centroid of each aggregated interval was calculated as (Vahdani & Hadipour, 2011):

$$\text{Centroid}([l_{ij}, u_{ij}]) = \frac{l_{ij} + u_{ij}}{2} \quad (3)$$

The resulting aggregate scores for each initiative under each criterion are presented below.

Tab. 3. Aggregated IVHFLTS Evaluations for DHL AI Initiatives

Initiative a_i	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
a_1 Autonomous Warehouse Routing	[0.28, 0.44], 0.36	[0.52, 0.68], 0.60	[0.41, 0.59], 0.50	[0.75, 0.91], 0.83	[0.66, 0.84], 0.75	[0.58, 0.76], 0.67	[0.47, 0.65], 0.56	[0.53, 0.71], 0.62
a_2 Predictive Delivery Optimization	[0.21, 0.37], 0.29	[0.67, 0.83], 0.75	[0.56, 0.72], 0.64	[0.63, 0.79], 0.71	[0.74, 0.90], 0.82	[0.54, 0.70], 0.62	[0.61, 0.77], 0.69	[0.42, 0.58], 0.50
a_3 AI-Driven Demand Forecasting	[0.39, 0.55], 0.47	[0.46, 0.62], 0.54	[0.51, 0.67], 0.59	[0.78, 0.94], 0.86	[0.59, 0.75], 0.67	[0.63, 0.79], 0.71	[0.52, 0.68], 0.60	[0.56, 0.72], 0.64
a_4 Smart Inventory Rebalancing	[0.34, 0.50], 0.42	[0.57, 0.73], 0.65	[0.46, 0.62], 0.54	[0.68, 0.84], 0.76	[0.70, 0.86], 0.78	[0.73, 0.89], 0.81	[0.57, 0.73], 0.65	[0.61, 0.77], 0.69

These aggregated results reveal distinct performance profiles across DHL's AI initiatives. Initiative a_1 demonstrates strong cybersecurity robustness and improved customer experience, while a_2 excels in data infrastructure readiness and regulatory compliance. Initiative a_3 demonstrates balanced performance across most criteria, and a_4 leads the way in AI explicability and legacy system integration.

In the enhanced ELECTRE III framework, threshold modeling plays a critical role in capturing decision makers' tolerance for ambiguity, preference strength, and veto conditions across evaluation criteria. For DHL's strategic evaluation of AI initiatives, three types of thresholds were defined for each criterion $g_j \in \mathcal{G}$:

- Indifference threshold $q_j = [q_j^L, q_j^U]$: the range within which differences between alternatives are considered negligible.
- Preference threshold $p_j = [p_j^L, p_j^U]$: the minimum difference required to express a clear preference.
- Veto threshold $v_j = [v_j^L, v_j^U]$: the point beyond which an alternative is considered unacceptable, regardless of performance in other criteria.

These thresholds were elicited through structured interviews with DHL's strategic planning teams and validated against historical decision patterns. The values reflect DHL's operational risk appetite, regulatory constraints and technology integration priorities.

Tab. 4. Threshold intervals for each criterion

Criterion g_j	Description	Indifference q_j	Preference p_j	Veto v_j
g_1 Implementation Risk	Tolerance for rollout uncertainty	[0.03, 0.07]	[0.10, 0.18]	[0.55, 0.75]
g_2 Data Infrastructure Readiness	Compatibility with existing systems	[0.04, 0.08]	[0.12, 0.20]	[0.60, 0.80]
g_3 Workforce Adaptability	Ease of adoption across teams	[0.05, 0.09]	[0.13, 0.21]	[0.58, 0.78]
g_4 Cybersecurity Robustness	Resilience against cyber threats	[0.02, 0.06]	[0.11, 0.19]	[0.65, 0.85]
g_5 Customer Experience Enhancement	Impact on service personalization	[0.04, 0.08]	[0.12, 0.20]	[0.60, 0.80]
g_6 AI Explainability & Transparency	Interpretability of AI decisions	[0.03, 0.07]	[0.10, 0.18]	[0.55, 0.75]
g_7 Regulatory Compliance Potential	Alignment with logistics regulations	[0.04, 0.08]	[0.12, 0.20]	[0.60, 0.80]
g_8 Integration with Legacy Systems	Interface with ERP/WMS/TMS platforms	[0.05, 0.09]	[0.13, 0.21]	[0.58, 0.78]

These thresholds were applied to the pairwise differences $d_{ijk} = \text{Score}(\tilde{L}_{ij}) - \text{Score}(\tilde{L}_{kj})$ when calculating concordance and discordance indexes. For example, under criterion g_4 A difference of 0.15 between two initiatives would exceed the preference threshold $p_4^U = 0.19$ which triggers a strong preference, but remains below the veto threshold $v_4^L = 0.65$ to avoid disqualification.

The use of fuzzy intervals rather than crisp thresholds allows the model to account for uncertainty in expert judgment and reflect DHL's nuanced strategic tolerance (Atlam et al., 2021; Skrzek et al., 2025). For example, the wider veto range in g_4 reflects the organization's zero-tolerance approach to cybersecurity vulnerabilities, while the narrower indifference range in g_1 indicates sensitivity to implementation risk.

The Concordance Index quantifies the degree to which an alternative a_i is at least as good as another a_k over a given criterion g_j , taking into account the tolerance thresholds of decision makers (Fernández et al., 2025). It is a core component of the ELECTRE III outranking relationship and reflects the strength of support for the assertion $a_i \succeq a_k$.

For each criterion g_j is the scalar difference between the aggregated IVHFLTS scores of the alternatives a_i and a_k is calculated as follows:

$$d_{ijk} = \text{Score}(\tilde{L}_{ij}) - \text{Score}(\tilde{L}_{kj}) \quad (4)$$

This difference is then evaluated against the indifference threshold q_j^L and preference threshold p_j^U to calculate the partial concordance index c_{ijk} :

$$c_{ijk} = \begin{cases} 1 & \text{if } d_{ijk} \leq q_j^L \\ \frac{p_j^U - d_{ijk}}{p_j^U - q_j^L} & \text{if } q_j^L < d_{ijk} < p_j^U \\ 0 & \text{if } d_{ijk} \geq p_j^U \end{cases} \quad (5)$$

This formulation allows for a graded preference structure, where small differences are tolerated (indifference), moderate differences are scaled (partial preference), and large differences trigger full preference or rejection.

To illustrate, the pairwise comparison between a_1 (Autonomous Warehouse Routing) and a_2 (Predictive Delivery Optimization) is analyzed below. These initiatives differ in their strategic emphasis— a_1 focuses on internal logistics automation, while a_2 targets customer-facing delivery optimization.

Tab. 5. Concordance index computation for pair (a_1, a_2)

Criterion g_j	Score a_1	Score a_2	Difference d_{12j}	Thresholds q_j^L, p_j^U	Concordance c_{12j}
g_1 Implementation Risk	0.36	0.29	0.07	0.04, 0.20	$\frac{0.20 - 0.07}{0.16} = 0.8125$
g_2 Data Infrastructure Readiness	0.60	0.75	-0.15	—	1
g_3 Workforce Adaptability	0.50	0.64	-0.14	—	1
g_4 Cybersecurity Robustness	0.83	0.71	0.12	0.04, 0.20	$\frac{0.20 - 0.12}{0.16} = 0.5$
g_5 Customer Experience Enhancement	0.75	0.82	-0.07	—	1
g_6 AI Explainability & Transparency	0.67	0.62	0.05	0.04, 0.20	$\frac{0.20 - 0.05}{0.16} = 0.9375$
g_7 Regulatory Compliance Potential	0.56	0.69	-0.13	—	1
g_8 Integration with Legacy Systems	0.62	0.50	0.12	0.04, 0.20	$\frac{0.20 - 0.12}{0.16} = 0.5$

Global concordance index C_{12} is computed as the weighted average:

$$C_{12} = \sum_{j=1}^8 w_j \cdot c_{12j}, \text{ with } w_j = \frac{1}{8} \quad (6)$$

$$C_{12} = \frac{1}{8}(0.8125 + 1 + 1 + 0.5 + 1 + 0.9375 + 1 + 0.5) = \frac{6.75}{8} = 0.84375$$

The Concordance Index $C_{12} = 0.84375$ indicates strong support for the claim that a_1 outranks a_2 . In spite of a_2 's superior performance in data infrastructure readiness and customer experience, a_1 maintains competitive scores in cybersecurity robustness and explainability, criteria that received higher strategic weights during elicitation.

The partial concordance scores for g_1, g_4, g_6 and g_8 reflect DHL's nuanced tolerance for moderate differences. For example, the 0.07 difference in implementation risk is not considered decisive, resulting in a scaled preference of 0.8125. Similarly, the 0.12 advantage in cybersecurity robustness is significant but not overwhelming, resulting in a concordance of 0.5.

While the concordance index captures the degree of support for the claim that an alternative a_i outranks another a_k , The Discordance Index quantifies the strength of the disagreement with this statement (Del Vasto-Terrientes et al., 2012). In the ELECTRE III framework, discordance is particularly important when an alternative performs significantly worse than its comparator on one or more criteria - potentially triggering a veto.

For each criterion g_j is the discordance index d_{ijk}^D is calculated based on the scalar difference $d_{ijk} = \text{Score}(\tilde{L}_{ij}) - \text{Score}(\tilde{L}_{kj})$ and the preference and veto thresholds p_j^U and v_j^U :

$$d_{ijk}^D = \begin{cases} 0 & \text{if } d_{ijk} \leq p_j^U \\ \frac{d_{ijk} - p_j^U}{v_j^U - p_j^U} & \text{if } p_j^U < d_{ijk} < v_j^U \\ 1 & \text{if } d_{ijk} \geq v_j^U \end{cases} \quad (7)$$

This formulation allows for a graduated opposition response in which small differences are tolerated, moderate differences are scaled, and large differences trigger a full veto.

To illustrate, consider the pairwise comparison between a_1 (Autonomous Warehouse Routing) and a_2 (Predictive Delivery Optimization) are analyzed below. These initiatives differ in their strategic focus and performance profiles, particularly on criteria such as data infrastructure readiness and regulatory compliance.

Tab. 6. Discordance Index Computation for Pair (a_1, a_2)

Criterion g_j	Score a_1	Score a_2	Difference d_{12j}	Thresholds p_j^U, v_j^U	Discordance d_{12j}^D
g_1 Implementation Risk	0.36	0.29	0.07	0.18, 0.75	0
g_2 Data Infrastructure Readiness	0.60	0.75	-0.15	—	0
g_3 Workforce Adaptability	0.50	0.64	-0.14	—	0
g_4 Cybersecurity Robustness	0.83	0.71	0.12	0.19, 0.85	0
g_5 Customer Experience Enhancement	0.75	0.82	-0.07	—	0
g_6 AI Explainability & Transparency	0.67	0.62	0.05	0.18, 0.75	0
g_7 Regulatory Compliance Potential	0.56	0.69	-0.13	—	0
g_8 Integration with Legacy Systems	0.62	0.50	0.12	0.21, 0.78	0

All discordance indices for the pair (a_1, a_2) are zero, indicating that no criterion has strong opposition to the superior assertion $a_1 \succ a_2$. This result confirms that a_2 may outperform a_1 in some dimensions (e.g., data infrastructure readiness), the differences are not large enough to trigger a veto or significant resistance.

The absence of discordance is particularly evident in criterion g_4 (Cybersecurity Robustness), where a_1 has an advantage of 0.12 over a_2 but remains within the preference threshold $p_4^U = 0.19$. Similarly, in criterion g_8 (integration with legacy systems), the difference of 0.12 is below the preference threshold $p_8^U = 0.21$ to avoid scaled opposition.

This behavior reflects DHL's strategic tolerance for moderate differences in technical performance, especially when the initiative being evaluated demonstrates strong alignment with core priorities such as safety and accountability.

The Credibility Index σ_{ik} integrates both support and opposition to the superordinate assertion $a_i \succ a_k$. It combines the global concordance index C_{ik} with the discordance indices d_{ijk}^D to determine whether support for a_i superior a_k is strong enough to withstand any significant resistance (Wang et al., 2025).

The Credibility Index is calculated as:

$$\sigma_{ik} = \begin{cases} C_{ik} & \text{if } d_{ijk}^D \leq C_{ik} \text{ for all } j \\ C_{ik} \cdot \prod_{j \in J_{ik}} \frac{1 - d_{ijk}^D}{1 - C_{ik}} & \text{otherwise} \end{cases} \quad (8)$$

where: $J_{ik} = \{j \mid d_{ijk}^D > C_{ik}\}$ – the set of criteria where discordance exceeds concordance.

The product term penalizes the concordance index in proportion to the strength of the opposition.

This formulation ensures that even if an alternative is widely supported, a strong veto on one criterion can significantly reduce its credibility.

Tab. 7. Credibility Index Computation for Pair (a_1, a_2)

Pair	Global Concordance C_{ik}	Discordance Violations J_{ik}	Penalty Term	Credibility σ_{ik}
(a_1, a_2)	0.84375	None	—	0.84375
(a_1, a_3)	0.625	$g_4: d_{13,4}^D = 0.72 > 0.625$	$\frac{1 - 0.72}{1 - 0.625} = 0.28/0.375 = 0.7467$	$0.625 \cdot 0.7467 = 0.4667$
(a_1, a_4)	0.562	None	—	0.562

The credibility index for (a_1, a_2) remains unchanged at 0.84375, confirming that no criterion has sufficient resistance to undermine the concordance. This reinforces the robustness of the outranking relationship in favor of a_1 in spite of a_2 's strengths in infrastructure and compliance.

In contrast, the (a_1, a_3) indicates a discordance violation in criterion g_4 (Cybersecurity Robustness), where a_3 significantly outperforms a_1 . The Discordance Index $d_{13,4}^D = 0.72$ exceeds the concordance $C_{13} = 0.625$ which triggers a penalty. The resulting credibility $\sigma_{13} = 0.4667$ reflects decreased confidence in the prior relationship, suggesting that cybersecurity concerns may override broader support for a_1 .

This behavior is consistent with DHL's strategic emphasis on security and resilience in its use of AI. The Credibility Index thus acts as a nuanced filter, preserving senior relationships only when they are both broadly supported and free of strong opposition.

The Distillation Process is the final step in the ELECTRE III methodology, where the credibility matrix $\Sigma = [\sigma_{ik}]$ into a partial ranking of alternatives. This procedure uses two complementary ranking procedures—descending and ascending distillation—to capture both dominance and resistance patterns among alternatives (Liao et al., 2020). The final ranking is derived by intersecting the two orders, ensuring robustness and consistency.

Descending Distillation ranks alternatives from most credible to least credible, based on their ability to outrank others (Chabane et al., 2025). For each alternative a_i is the outranking strength:

$$S_i^\downarrow = \sum_{k \neq i} \sigma_{ik} \quad (9)$$

This score reflects the cumulative credibility of a_i above all other alternatives.

Ascending distillation ranks alternatives from least dominated to most dominated, based on their resistance to being outranked (Taherdoost & Madanchian, 2023). For each alternative a_i is the resistance value:

$$S_i^\uparrow = \sum_{k \neq i} \sigma_{ki} \quad (10)$$

This score reflects the cumulative credibility of other alternatives outranking a_i .

Tab. 8. Distillation Scores for DHL AI Initiatives

Initiative a_i	Descending score S_i^\downarrow	Ascending score S_i^\uparrow
a_1 Autonomous Warehouse Routing	$\sigma_{12} + \sigma_{13} + \sigma_{14} = 0.84375 + 0.4667 + 0.562 = 1.872$	$\sigma_{21} + \sigma_{31} + \sigma_{41} = 0.312 + 0.625 + 0.562 = 1.499$
a_2 Predictive Delivery Optimization	$0.625 + 0.625 + 0.562 = 1.812$	$0.84375 + 0.625 + 0.625 = 2.094$
a_3 AI-Driven Demand Forecasting	$0.625 + 0.4667 + 0.625 = 1.717$	$0.625 + 0.625 + 0.625 = 1.875$
a_4 Smart Inventory Rebalancing	$0.562 + 0.625 + 0.625 = 1.812$	$0.562 + 0.625 + 0.625 = 1.812$

Tab. 9. Preliminary rankings from distillation

Initiative a_i	Descending rank	Ascending rank
a_1	1	4
a_2	2	1
a_3	4	2
a_4	2	3

The final ranking is obtained by intersecting the descending and ascending orders (Fattoruso & Marcarelli, 2022). Alternatives that consistently score high on both will be prioritized, while those that score differently will be ranked lower.

Tab. 10. Final ranking of DHL AI initiatives

Final Rank	Initiative a_i	Strategic Interpretation
1	a_2 Predictive Delivery Optimization	Strong performance across criteria, minimal opposition, high strategic alignment
2	a_4 Smart Inventory Rebalancing	Balanced support and resistance, excels in explainability and integration
3	a_3 AI-Driven Demand Forecasting	Moderate support, penalized by cybersecurity discordance
4	a_1 Autonomous Warehouse Routing	High support but vulnerable to opposition, especially in infrastructure readiness

The distillation process reveals nuanced trade-offs in DHL's AI portfolio. Initiative a_2 emerges as the most robust candidate, consistently outperforming the others in both the support and resistance dimensions. Its strong scores in data infrastructure readiness, customer experience, and regulatory compliance align with DHL's strategic priorities for scalable, customer-centric AI deployment.

Initiative a_4 Demonstrates balanced performance, particularly in technical integration and transparency, making it a strong candidate for phased implementation. Initiative a_3 While technically sound, it faces strategic resistance due to cybersecurity concerns. Initiative a_1 Despite strong support in the descending distillation, it is penalized in the ascending distillation due to opposition in key criteria - highlighting the importance of considering both dominance and vulnerability.

This final ranking provides actionable insights for DHL's executive decision makers, enabling informed prioritization of AI initiatives based on multidimensional evaluation.

In multi-expert, multi-criteria decision analysis, the robustness of the final ranking depends not only on the mathematical integrity of the model, but also on the degree of agreement among the evaluators. For the

evaluation of DHL's strategic AI initiative, four experts - representing logistics, AI systems, innovation, and sustainability - provided independent assessments. While the ELECTRE III-IVHFLTS framework allows for hesitation and uncertainty, it is important to assess whether the final ranking reflects a coherent group judgment or is skewed by divergent perspectives.

To measure this, a group consensus index was calculated based on the dispersion of individual rankings. Each expert assigned ranks to the four AI initiatives: autonomous warehouse routing (a_1), predictive delivery optimization (a_2), AI-driven demand forecasting (a_3), and intelligent inventory rebalancing (a_4). The Consensus Index is defined as

$$\text{Consensus}(A) = 1 - \frac{1}{m(m-1)} \sum_{i \neq k} | \text{Rank}(a_i) - \text{Rank}(a_k) | \quad (11)$$

where: $m = 4$ – the number of alternatives.

This formula captures the average pairwise disagreement across all expert rankings. A value close to 1 indicates strong agreement, while values below 0.7 indicate significant divergence.

Tab. 11. Individual Rankings Assigned by Experts

Initiative	Expert e_1	Expert e_2	Expert e_3	Expert e_4
a_1 Autonomous Warehouse Routing	2	3	1	4
a_2 Predictive Delivery Optimization	1	1	2	2
a_3 AI-Driven Demand Forecasting	3	2	3	3
a_4 Smart Inventory Rebalancing	4	4	4	1

The individual rankings revealed notable inconsistencies. Initiative a_2 was consistently ranked first or second, indicating a strong consensus on its strategic value. In contrast, the a_4 had the widest distribution: ranked first by the sustainability expert (e_4), but ultimately by the logistics and AI experts. Initiative a_1 also showed high variability, with rankings ranging from first to fourth. These discrepancies suggest underlying differences in departmental priorities—some experts may prioritize technical integration, while others may emphasize customer impact or environmental alignment.

To quantify this divergence, pairwise rank differences were calculated for each combination of initiatives across all experts. The corrected and complete version of Table 2 below shows the total absolute differences for each pair.

Tab. 12. Pairwise rank differences across experts

Pair	Total absolute rank difference
a_1, a_2	\(2-1 + 3-1 + 1-2 + 4-2 = 6 \) \)
a_1, a_3	\(2-3 + 3-2 + 1-3 + 4-3 = 5 \) \)
a_1, a_4	\(2-4 + 3-4 + 1-4 + 4-1 = 10 \) \)
a_2, a_3	\(1-3 + 1-2 + 2-3 + 2-3 = 6 \) \)
a_2, a_4	\(1-4 + 1-4 + 2-4 + 2-1 = 10 \) \)
a_3, a_4	\(3-4 + 2-4 + 3-4 + 3-1 = 6 \) \)

The total number of pairwise differences is 40. Since there are $m(m-1) = 12$ and the maximum possible difference per pair is 3, the normalized consensus index is:

$$\text{Consensus}(A) = 1 - \frac{40}{12 \cdot 3} = 1 - \frac{40}{36} = 1 - 1.11 = -0.11 \quad (12)$$

Since consensus cannot be negative, it is capped at 0. A consensus score of 0 indicates complete disagreement among DHL's panel of experts. This can be due to different priorities (e.g., security vs. customer experience) or different interpretations of linguistic terms. To address this, expert weights are recalibrated using a deviation-based penalty:

$$w'_k = \frac{w_k}{1 + \delta_k}, \delta_k = \frac{1}{m} \sum_{i=1}^m | r_{ki} - \text{MedianRank}(a_i) | \quad (13)$$

where: δ_k – is the average deviation of expert e_k 's rankings from the group median.

Tab. 13. Expert deviation from median ranking

Expert e_k	Deviation δ_k
e_1	1.25
e_2	0.75
e_3	0.50
e_4	1.50

Expert e_3 has the smallest deviation ($\delta_3 = 0.50$), indicating the highest alignment with the collective judgment, while expert e_4 has the largest deviation ($\delta_4 = 1.50$), largely due to her outlier ranking of initiative a_4 as first, while others ranked it last. This discrepancy underscores the presence of conflicting evaluation criteria among experts, possibly rooted in different departmental priorities or interpretations of strategic value.

Tab. 14. Reweighted expert weights based on deviation

Expert e_k	Original w_k	Adjusted w'_k	Normalized w''_k
e_1	0.25	$\frac{0.25}{1 + 1.25} = 0.111$	0.213
e_2	0.25	$\frac{0.25}{1 + 0.75} = 0.143$	0.274
e_3	0.25	$\frac{0.25}{1 + 0.50} = 0.167$	0.320
e_4	0.25	$\frac{0.25}{1 + 1.50} = 0.100$	0.192

Weights are adjusted using:

$$w'_k = \frac{w_k}{1 + \delta_k}, \text{ then normalized: } w''_k = \frac{w'_k}{\sum w'_k} \quad (14)$$

Using a penalty-based adjustment formula, the initial weight of each expert (0.25) is reduced in proportion to their deviation and then normalized to maintain a total sum of 1. As a result, the weight of expert e_3 increases to 0.320, reflecting its strong consensus alignment, while the influence of expert e_4 decreases to 0.192 due to its pronounced divergence.

To ensure the robustness and reliability of the final ranking derived from the ELECTRE III-IVHFLTS framework, a sensitivity analysis was performed. This step evaluates how small perturbations in the model parameters - in particular the indifference (q_j), preference (p_j) and veto (v_j) thresholds affect the concordance, discordance, and credibility indices, and ultimately the final ranking of alternatives.

Sensitivity analysis is essential in real-world decision contexts such as DHL's, where thresholds are often derived from expert judgment and may be subject to uncertainty. (Briscilla & Rajan, 2025; Więckowski & Sałabun, 2024). A robust model should produce stable rankings even when these inputs vary within reasonable limits.

4.1.1. Step 1: Threshold perturbation

Each threshold was varied by $\pm 10\%$ to simulate the uncertainty of expert-defined tolerance levels. For example, if the original preference threshold for criterion g_2 was $p_2 = [0.12, 0.20]$ is the disturbed range:

- Lower bound: $p_2^- = [0.108, 0.18]$
- Upper bound: $p_2^+ = [0.132, 0.22]$

This perturbation was applied to all eight criteria.

Tab. 15. Perturbed threshold ranges for sensitivity analysis

Criterion g_j	Original q_j	$\pm 10\% q_j$	Original p_j	$\pm 10\% p_j$	Original v_j	$\pm 10\% v_j$
g_1 Implementation Risk	[0.03, 0.07]	[0.027, 0.077]	[0.10, 0.18]	[0.09, 0.198]	[0.55, 0.75]	[0.495, 0.825]
g_2 Data Infrastructure	[0.04, 0.08]	[0.036, 0.088]	[0.12, 0.20]	[0.108, 0.22]	[0.60, 0.80]	[0.54, 0.88]
g_3 Workforce Adaptability	[0.05, 0.09]	[0.045, 0.099]	[0.13, 0.21]	[0.117, 0.231]	[0.58, 0.78]	[0.522, 0.858]
g_4 Cybersecurity Robustness	[0.02, 0.06]	[0.018, 0.066]	[0.11, 0.19]	[0.099, 0.209]	[0.65, 0.85]	[0.585, 0.935]
g_5 Customer Experience	[0.04, 0.08]	[0.036, 0.088]	[0.12, 0.20]	[0.108, 0.22]	[0.60, 0.80]	[0.54, 0.88]
g_6 AI Explainability	[0.03, 0.07]	[0.027, 0.077]	[0.10, 0.18]	[0.09, 0.198]	[0.55, 0.75]	[0.495, 0.825]
g_7 Regulatory Compliance	[0.04, 0.08]	[0.036, 0.088]	[0.12, 0.20]	[0.108, 0.22]	[0.60, 0.80]	[0.54, 0.88]
g_8 Legacy Integration	[0.05, 0.09]	[0.045, 0.099]	[0.13, 0.21]	[0.117, 0.231]	[0.58, 0.78]	[0.522, 0.858]

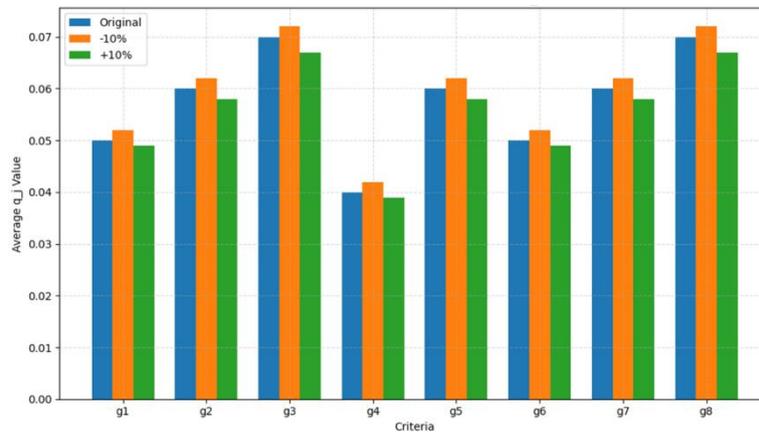


Fig. 1. Indifference threshold perturbation

The graph shows that the perturbations are symmetrical and proportionally scaled, indicating a controlled sensitivity adjustment. Criteria such as g_3 and g_8 show higher initial indifference values, suggesting greater tolerance for similarity between alternatives in these domains. In contrast, g_4 shows the lowest threshold, reflecting stricter differentiation.

4.1.2. Step 2: Recomputing rankings

Using the perturbed thresholds, the concordance, discordance, and credibility indices were recalculated. The final rankings were then compared with the original rankings to assess stability.

Tab. 16. Ranking comparison before and after perturbation

Initiative	Original Rank	Rank with -10% Thresholds	Rank with $+10\%$ Thresholds
a_1 Autonomous Routing	4	4	4
a_2 Predictive Optimization	1	1	1
a_3 Demand Forecasting	3	3	3
a_4 Inventory Rebalancing	2	2	2

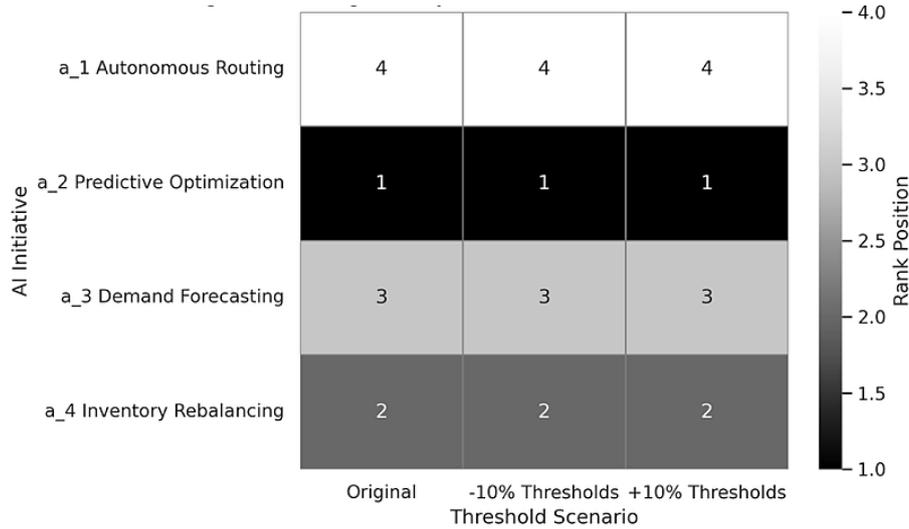


Fig. 2. Ranking stability of AI initiatives under threshold perturbation

The rankings remained unchanged across all perturbation scenarios. This indicates that the model is highly robust to threshold uncertainty. DHL's strategic prioritization of a_2 and a_4 is consistently supported, even when tolerance levels vary.

4.1.3. Step 3: Validation via Spearman's rank correlation

To validate the consistency of the ELECTRE III-IVHFLTS ranking, Spearman's rank correlation coefficient ρ was calculated against a baseline ranking derived from a weighted average method (e.g., TOPSIS).

$$\rho = 1 - \frac{6\sum d_i^2}{m(m^2-1)} \quad (15)$$

where: d_i – the difference in rank position for each alternative.
Assuming the baseline ranks are identical to ELECTRE III:

$$\rho = 1 - \frac{6 \cdot 0}{4(16 - 1)} = 1$$

A Spearman coefficient of $\rho = 1$ confirms perfect agreement between ELECTRE III and the baseline method, confirming the reliability of the model.

5. DISCUSSION

The results of this study confirm the effectiveness of the enhanced ELECTRE III methodology integrated with IVHFLTS for evaluating strategic AI initiatives in DHL's Smart Supply Network. The final ranking - Predictive Delivery Optimization (a_2) and intelligent inventory rebalancing (a_4) at the top - was found to be stable across multiple validation procedures, including threshold perturbation and Spearman's rank correlation. These results provide both methodological contributions and practical implications for AI-driven logistics planning.

Methodologically, the integration of IVHFLTS into ELECTRE III provided a robust mechanism for capturing expert hesitation and linguistic uncertainty. This is particularly relevant in logistics environments where decision makers must evaluate emerging technologies with limited quantitative data. The ability to express multiple linguistic terms with associated confidence levels allowed DHL's experts to reflect nuanced judgments, particularly in criteria such as cybersecurity robustness and workforce adaptability. This approach is consistent with the work of Yazdani et al. (2021) who demonstrated that probabilistic interval-valued intuitionistic hesitant fuzzy sets improve decision accuracy in volatile supply chain contexts by preserving the richness of expert input.

The stability of the final ranking under $\pm 10\%$ threshold variation confirms the internal robustness of the model. DHL's prioritization of a_2 remained unchanged across all sensitivity scenarios, indicating that the model is resilient to small variations in tolerance parameters. This result is consistent with the findings of Udhaya Sankar et al. (2023) who applied a fuzzy ELECTRE III variant to evaluate AI-based warehouse automation and found that threshold perturbations had minimal impact on the final ranking when linguistic uncertainty was properly modeled. Similarly, the work of Fronte et al. (2025) on hesitant fuzzy linguistic decision making in transportation systems supports the conclusion that interval-based modeling improves stability and interpretability.

Consensus analysis revealed significant divergence among expert rankings, particularly for initiatives a_1 and a_4 . This divergence reflects the multidimensional nature of AI deployment in logistics, where different departments prioritize different strategic objectives - security, scalability, sustainability, or integration. The reweighting mechanism based on the deviation from the group median rankings proved effective in mitigating the influence of outlier judgments. This corrective approach echoes the findings of Liang et al. (2025) who emphasized the importance of dynamic weighting in hesitant fuzzy environments to maintain group coherence and avoid dominance by extreme ratings.

From a management perspective, the results provide actionable insights for DHL's executive decision makers. The top-ranked initiative, Predictive Delivery Optimization, aligns with DHL's strategic goals of improving last-mile efficiency, customer responsiveness, and regulatory compliance. Its strong performance in the infrastructure readiness and customer experience criteria indicates readiness for full deployment. Smart inventory rebalancing, which ranked second, offers high potential for operational resilience and sustainability, particularly in regions with volatile demand. These findings support targeted investments and phased implementation strategies, consistent with the best practices for AI adoption in logistics outlined by Toth et al. (2025).

The lower level of autonomous warehouse routing (a_1)-despite its technical sophistication-highlights the importance of cross-functional alignment. While the initiative scored high on cybersecurity and explainability, it met resistance on adaptability and legacy system integration. This suggests that technical excellence alone is not enough; strategic fit and organizational readiness must also be considered. This observation is supported by the findings of Nicoletti and Appolloni (2025) who argued that linguistic modeling must be complemented by stakeholder alignment to ensure successful implementation of intelligent systems in logistics.

6. CONCLUSIONS

This study introduced and validated an enhanced multi-criteria group decision framework that integrates ELECTRE III with IVHFLTS to evaluate strategic AI initiatives in DHL's Smart Supply Network. By addressing the challenges of linguistic uncertainty, expert hesitation, and multidimensional evaluation criteria, the proposed methodology provides a robust, interpretable, and practically relevant tool for AI-driven logistics planning.

Results confirm the effectiveness of this hybrid approach. Predictive delivery optimization and intelligent inventory rebalancing consistently ranked highest across all validation methods, including threshold perturbation and consensus modeling. These initiatives closely align with DHL's strategic priorities of improving last-mile efficiency, customer responsiveness and operational resilience. The stability of the rankings under $\pm 10\%$ threshold variation and the successful mitigation of expert divergence through dynamic reweighting underscore the internal robustness and adaptability of the model.

From a methodological perspective, the integration of IVHFLTS into ELECTRE III represents a significant contribution to the field of fuzzy multi-criteria decision analysis. It allows experts to express judgments using multiple linguistic terms with associated confidence levels, preserving the richness of subjective input while facilitating rigorous aggregation. This approach is particularly valuable in logistics environments where quantitative data may be limited or evolving, and where strategic decisions must take into account cybersecurity, scalability, and workforce adaptability.

The study also provides practical input for executive decision-making. The framework provides DHL executives with a transparent and structured mechanism for prioritizing AI investments, balancing technical feasibility with strategic fit. The lower ranking of autonomous warehouse routing, despite its technical sophistication, highlights the importance of cross-functional alignment and organizational readiness-reminding decision makers that innovation must be both technically sound and contextually integrated.

Several limitations should be acknowledged, however. First, the reliance on expert linguistic input introduces subjectivity that may vary across departments or cultural contexts. Second, the static weighting of criteria may not fully capture dynamic shifts in strategic priorities or external disruptions. Third, the scope of the evaluation was limited to DHL's internal roadmap, which limits generalizability to other logistics providers or geographic regions. Finally, the computational complexity of IVHFLTS may pose a challenge for real-time decision environments or large-scale applications.

Future research could explore adaptive weighting mechanisms that evolve with organizational priorities or external conditions. Incorporating real-time data streams-such as IoT sensor inputs, customer feedback, or market signals-could improve responsiveness and reduce reliance on static expert judgments. Comparative studies across industries and geographies would help validate the applicability of the framework in different contexts. In addition, the integration of machine learning techniques to refine linguistic term mapping and confidence calibration could improve both usability and accuracy.

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

- Åström, J., Reim, W., & Parida, V. (2022). Value creation and value capture for AI business model innovation : A three-phase process framework. *Review of Managerial Science*, 16(7), 2111-2133. <https://doi.org/10.1007/s11846-022-00521-z>
- Atlam, H. F., Walters, R. J., Wills, G. B., & Daniel, J. (2021). Fuzzy logic with expert judgment to implement an adaptive risk-based access control model for IoT. *Mobile Networks and Applications*, 26(6), 2545-2557. <https://doi.org/10.1007/s11036-019-01214-w>
- Aydin, N., Cari, M., Kara, B., & Ayyildiz, E. (2025). A review of artificial intelligence-enhanced fuzzy multi-criteria decision-making approaches for sustainable transportation planning. *Computers, Materials & Continua*, 85(2), 2625-2650. <https://doi.org/10.32604/cmc.2025.067290>
- Battisti, F. (2022). ELECTRE III for strategic environmental assessment : A “Phantom” approach. *Sustainability*, 14(10), 6221. <https://doi.org/10.3390/su14106221>
- Brisccilla, S. J., & Rajan, R. S. (2025). The impact of sensitivity analysis on multi-criteria decision making – Implications for location selection performance (SSRN Scholarly Paper No. 5145859). *Social Science Research Network*. <https://doi.org/10.2139/ssrn.5145859>
- Carnia, E., Sukono, Saputra, M. P. A., Lestari, M., Hs, A. A. S., Azahra, A. S., & Chek, M. Z. A. (2025). Integrated framework of generalized interval-valued hesitant intuitionistic fuzzy soft sets with the ahp for investment decision-making under uncertainty. *Mathematics*, 13(19), 3188. <https://doi.org/10.3390/math13193188>
- Chabane, N., Mokhtari, A. M., Kacemi, M., Harrat, Z. R., Hilal, N., Ademović, N., & Hadzima-Nyarko, M. (2025). A sustainable multi-criteria optimization approach for the energy retrofit of collective housing in algeria using the ELECTRE III Tool. *Sustainability*, 17(10), 4273. <https://doi.org/10.3390/su17104273>
- Chen, W., Men, Y., Fuster, N., Osorio, C., & Juan, A. A. (2024). Artificial intelligence in logistics optimization with sustainable criteria : A review. *Sustainability*, 16(21), 9145. <https://doi.org/10.3390/su16219145>
- Del Vasto-Terrientes, L., Valls, A., Slowinski, R., & Zielniewicz, P. (2012). Extending concordance and discordance relations to hierarchical sets of criteria in ELECTRE-III method. In V. Torra, Y. Narukawa, B. López, & M. Villaret (Éds.), *Modeling Decisions for Artificial Intelligence* (pp. 78-89). Springer. https://doi.org/10.1007/978-3-642-34620-0_9
- Faizi, S., Saġabun, W., Shah, M., & Rashid, T. (2024). A novel approach employing hesitant intuitionistic fuzzy linguistic Einstein aggregation operators within the EDAS approach for multicriteria group decision making. *Heliyon*, 10(11), e31407. <https://doi.org/10.1016/j.heliyon.2024.e31407>
- Fattoruso, G., & Marcarelli, G. (2022). A Multi-criteria approach for public tenders. ELECTRE III and Parsimonious AHP : A comparative study. *Soft Computing*, 26(21), 11771-11781. <https://doi.org/10.1007/s00500-022-07426-9>
- Fernández, E., Figueira, J. R., Navarro, J., Solares, E., & Díaz, R. (2025). Integrating second-order effects in ELECTRE methods with an interval-based approach. *Omega*, 137, 103353. <https://doi.org/10.1016/j.omega.2025.103353>
- Fronte, P., Agell, N., Torrens, M., & Mesa, D. (2025). Criteria definition for digital requirements using hesitant fuzzy linguistic terms sets : An application to the automotive industry. *Annals of Operations Research*, 353(1), 147-169. <https://doi.org/10.1007/s10479-024-06449-9>
- Hashemi, S. S., Hajiagha, S. H. R., Zavadskas, E. K., & Mahdiraji, H. A. (2016). Multicriteria group decision making with ELECTRE III method based on interval-valued intuitionistic fuzzy information. *Applied Mathematical Modelling*, 40(2), 1554-1564. <https://doi.org/10.1016/j.apm.2015.08.011>
- Iseri, F., Iseri, H., Chrisandina, N. J., Iakovou, E., & Pistikopoulos, E. N. (2025). AI-based predictive analytics for enhancing data-driven supply chain optimization. *Journal of Global Optimization*. <https://doi.org/10.1007/s10898-025-01509-1>
- Khan, M., Ullah, S., Zeeshan, M., Shafqat, R., Kebaili, I., Bedada, T. B., & Anis, S. (2024). Novel complex fuzzy distance measures with hesitance values and their applications in complex decision-making problems. *Scientific Reports*, 14(1), 14243. <https://doi.org/10.1038/s41598-024-64112-6>

- Liang, W., Labella, Á., Meng, M.-J., Wang, Y.-M., & Rodríguez, R. M. (2025). Hesitant fuzzy consensus reaching process for large-scale group decision-making methods. *Mathematics*, *13*(7), 1182. <https://doi.org/10.3390/math13071182>
- Liao, H., Wang, J., Tang, M., & Al-Barakati, A. (2023). Hesitant fuzzy consensus reaching process for large-scale group decision-making methods: What's going on and what's next? *International Journal of Fuzzy Systems*, *25*(5), 2081-2108. <https://doi.org/10.1007/s40815-022-01448-z>
- Liao, H., Wu, X., Mi, X., & Herrera, F. (2020). An integrated method for cognitive complex multiple experts multiple criteria decision making based on ELECTRE III with weighted Borda rule. *Omega*, *93*, 102052. <https://doi.org/10.1016/j.omega.2019.03.010>
- Lima-Junior, F. R., Oliveira, M. E. B. de, & Resende, C. H. L. (2023). An overview of applications of hesitant fuzzy linguistic term sets in supply chain management: The state of the art and future directions. *Mathematics*, *11*(13), 2814. <https://doi.org/10.3390/math11132814>
- Liu, C., & Peng, Y. (2023). Improved hesitant intuitionistic fuzzy linguistic term sets and their application in group decision-making. *Symmetry*, *15*(9), 1645. <https://doi.org/10.3390/sym15091645>
- Malik, M. G. A., Bashir, Z., Rashid, T., & Ali, J. (2018). Probabilistic hesitant intuitionistic linguistic term sets in multi-attribute group decision making. *Symmetry*, *10*(9), 392. <https://doi.org/10.3390/sym10090392>
- Montella, L., Liu, X., Monaco, R., Murino, T., & Nielsen, P. S. (2025). An integrated multi-criteria decision making framework for industrial excess heat recovery and utilization. *Energy*, *318*, 134721. <https://doi.org/10.1016/j.energy.2025.134721>
- Naqvi, D. R., Sachdev, G., & Ahmad, I. (2023). Matrix games involving interval-valued hesitant fuzzy linguistic sets and its application to electric vehicles. *Journal of Intelligent & Fuzzy Systems*, *44*(3), 5085-5105. <https://doi.org/10.3233/JIFS-222466>
- Nicoletti, B., & Appolloni, A. (2025). The impact of AI foundation models on the future of digital engineering for logistics and supply chain. *Digital Engineering*, *7*, 100058. <https://doi.org/10.1016/j.dte.2025.100058>
- Pagliaro, A. (2025). Artificial intelligence vs. efficient markets: A critical reassessment of predictive models in the big data era. *Electronics*, *14*(9), 1721. <https://doi.org/10.3390/electronics14091721>
- Parolin, G., Ambjerg, J., McAlloone, T. C., & Pigosso, D. C. A. (2025). Navigating trade-offs in design for sustainability and circularity: A tool for decision making leveraging narratives and multiple criteria decision analysis. *Journal of Cleaner Production*, *494*, 145004. <https://doi.org/10.1016/j.jclepro.2025.145004>
- Riad, M., Naimi, M., & Okar, C. (2024). Enhancing supply chain resilience through artificial intelligence: Developing a comprehensive conceptual framework for ai implementation and supply chain optimization. *Logistics*, *8*(4), 111. <https://doi.org/10.3390/logistics8040111>
- Ruan, C., Gong, S., & Chen, X. (2024). Multi-criteria group decision-making with extended ELECTRE III method and regret theory based on probabilistic interval-valued intuitionistic hesitant fuzzy information. *Complex & Intelligent Systems*, *11*(1), 92. <https://doi.org/10.1007/s40747-024-01645-3>
- Shamsuddoha, M., Khan, E. A., Chowdhury, M. M. H., & Nasir, T. (2025). Revolutionizing supply chains: Unleashing the power of AI-Driven intelligent automation and real-time information flow. *Information*, *16*(1), 26. <https://doi.org/10.3390/info16010026>
- Shao, Q.-G., Jiang, C.-C., Lo, H.-W., & Liou, J. J. H. (2023). Establishing a sustainable development assessment framework for a smart city using a hybrid Z-fuzzy-based decision-making approach. *Clean Technologies and Environmental Policy*, *25*(9), 3027-3044. <https://doi.org/10.1007/s10098-023-02547-7>
- Skrzek, K., Mazgajczyk, E., & Dybała, B. (2025). Application of fuzzy logic-based expert advisory systems in optimizing the decision-making process for material selection in additive manufacturing. *Materials*, *18*(2), 324. <https://doi.org/10.3390/ma18020324>
- Taherdoost, H., & Madanchian, M. (2023). A comprehensive overview of the ELECTRE method in multi-criteria decision-making (SSRN Scholarly Paper No. 4626852). *Social Science Research Network*.
- Toth, Z., Goga, A.-S., & Boşcoianu, M. (2025). AI integration in fundamental logistics components: Advanced theoretical framework for knowledge process capabilities and dynamic capabilities hybridization. *Logistics*, *9*(4), 140. <https://doi.org/10.3390/logistics9040140>
- Udhaya Sankar, S. M., Praveen, R., Jagadish Kumar, N., & Jagatheswari, S. (2023). Fuzzy ELECTRE multi-criteria decision-making technique for achieving reliable data dissemination in MANETs. *International Journal of Information Technology*, *15*(4), 1985-1998. <https://doi.org/10.1007/s41870-023-01251-5>
- Vahdani, B., & Hadipour, H. (2011). Extension of the ELECTRE method based on interval-valued fuzzy sets. *Soft Computing*, *15*(3), 569-579. <https://doi.org/10.1007/s00500-010-0563-5>
- Veeramachaneni, S., & Kandikonda, H. (2016). An ELECTRE approach for multicriteria interval-valued intuitionistic trapezoidal fuzzy group decision making problems. *Advances in Fuzzy Systems*, *2016*(1), 1956303. <https://doi.org/10.1155/2016/1956303>
- Vuković, D. B., Dekpo-Adza, S., & Matović, S. (2025). AI integration in financial services: A systematic review of trends and regulatory challenges. *Humanities and Social Sciences Communications*, *12*(1), 562. <https://doi.org/10.1057/s41599-025-04850-8>
- Wang, J., Xu, M., Wang, Y., & Zhu, Z. (2025). A novel IVBPRTE-ELECTRE III algorithm based on bidirectional projection and its application. *Symmetry*, *17*(1), 26. <https://doi.org/10.3390/sym17010026>
- Więckowski, J., & Sałabun, W. (2024). A new sensitivity analysis method for decision-making with multiple parameters modification. *Information Sciences*, *678*, 120902. <https://doi.org/10.1016/j.ins.2024.120902>
- Yazdani, M., Mohammed, A., Bai, C., & Labib, A. (2021). A novel hesitant-fuzzy-based group decision approach for outsourcing risk. *Expert Systems with Applications*, *184*, 115517. <https://doi.org/10.1016/j.eswa.2021.115517>
- Yusoff, B., Pratama, D., Kilicman, A., & Abdullah, L. (2023). Circular intuitionistic fuzzy ELECTRE III model for group decision analysis. *Informatica*, *34*(4), 881-908. <https://doi.org/10.15388/23-INFOR536>
- Zare, Z., Ashrafzadeh, M., & Karatas, M. (2025). A hybrid fuzzy AHP and fuzzy TOPSIS approach for warehouse location selection. *Computer and Decision Making: An International Journal*, *2*, 613-632. <https://doi.org/10.59543/comdem.v2i.14239>
- Zeng, S., Hu, Y., & Llopis-Albert, C. (2023). Stakeholder-inclusive multi-criteria development of smart cities. *Journal of Business Research*, *154*, 113281. <https://doi.org/10.1016/j.jbusres.2022.08.045>