



The AI-driven Decision-Making (AIDM) Framework: Integrating AHP and ChatGPT for Supplier Selection

Negar Sadeghi¹ · Mohammad Dehghani¹ · Nihan Kabadayi²

Received: 24 January 2025 / Accepted: 23 February 2026 / Published online: 12 March 2026
© The Author(s) 2026

Abstract

The integration of Artificial Intelligence (AI) into decision-making is transforming industry norms by enabling smarter, faster, and more reliable decisions. This paper presents a novel approach for supplier selection in manufacturing by combining the Analytic Hierarchy Process (AHP) with the Generative Pre-trained Transformer (GPT). Using GPT as virtual agents to mimic expert evaluations automates the supplier selection process, significantly improving efficiency. The study demonstrates the effectiveness of large language models as virtual experts in complex decision-making, highlighting AI's crucial role in strategic manufacturing operations. Comparing GPT's assessments with human expert judgments confirms the model's reliability and effectiveness in optimizing supplier selection. This approach enhances decision-making processes and emphasizes the broad utility of AI-centric methodologies in various industries.

Keywords Analytic Hierarchy Process · Large Language Models · AI-enabled Decision Making · Supplier Selection · Generative AI

1 Introduction

The advancement of Large Language Models (LLMs), a form of Generative AI (GenAI), has had significant impacts across diverse domains. Their widespread adoption is remarkable, with millions of users interacting daily (Pandey et al., 2024). Trained on massive and heterogeneous datasets, LLMs possess comprehensive knowledge that enables them to approximate the reasoning of subject-matter experts across many fields. Recent studies demonstrate the effectiveness of LLMs in medical decision-making (Zardasht Mahmud Ahamed et al., 2023), supplier selection (Wang & Wu, 2024), and business decision-making, such as drafting marketing content and brainstorming ideas (Chuma & Oliveira, 2023).

Given this potential, the integration of GenAI into Multi-Criteria Decision Making (MCDM) frameworks is particularly promising. A major challenge in conducting MCDM studies is the reliance on scarce and often difficult-to-access domain experts. Substituting these traditional experts with LLM-driven virtual experts can alleviate this barrier, enhanc-

✉ Mohammad Dehghani
m.dehghani@northeastern.edu

¹ Mechanical and Industrial Engineering Department, Northeastern University, Boston, USA

² Production Management Department, Business School of İstanbul University, İstanbul, Turkey

ing the scalability and responsiveness of the decision-making process. Moreover, GenAI's ability to rapidly synthesize and evaluate large datasets across multiple criteria fosters a more dynamic and adaptive decision environment. This capability supports robust, data-driven decision-making across industries and represents a meaningful integration of AI with human-like reasoning (Zuheros et al., 2024).

This study showcases this potential by applying the Analytic Hierarchy Process (AHP) to a supplier selection problem in a machinery manufacturing context. Outcomes obtained through the traditional expert-driven AHP are compared with those generated by LLM-based virtual experts. Accordingly, this paper makes the following contributions:

- Introduction of the AI-Driven Decision Making (AIDM) framework, which utilizes GenAI for multiple MCDM analysis tasks.
- Demonstration of AIDM's integration of AHP and ChatGPT, showcasing its effectiveness in supplier selection.
- An open and reusable experimentation platform is presented, enabling researchers and practitioners to replicate, adapt, and extend the AIDM framework for diverse decision-making applications.

Recent research suggests that LLMs are capable of producing structured and explainable judgments that resemble those of human experts. Despite these advances, most hybrid AI—MCDM studies continue to rely on conventional AI techniques such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), or Decision Trees (DTs) (Fallahpour et al., 2016, 2018; Kumar & Roy, 2010; Lakshmanpriya et al., 2013; Tang et al., 2013; Desheng, 2009). Limited attention has been given to the integration of LLMs into multi-criteria decision-making frameworks, and little is known about how their generated pairwise comparisons within AHP align with assessments from domain specialists in real-world applications such as supplier selection. Accordingly, this study addresses the following research question: Can generative AI serve as a reliable virtual expert in an AHP-based supplier-selection task, producing judgments that are consistent with those of human decision-makers?

2 Literature Review

Supplier selection is a critical component of supply-chain management because it influences cost, quality, lead time and risk. Early research on supplier evaluation relied on multi-criteria decision-making (MCDM) methods, such as AHP and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), to provide transparent and systematic comparisons of alternatives. These methods structure the problem into hierarchical criteria and use expert judgments to assign weights and ranks. Although MCDM techniques are easy to interpret, they may assume uniform weight distribution across criteria and lack dimensional reduction capabilities, making them less effective when large numbers of criteria are involved. As modern supply chains generate large, complex data sets with many supplier attributes, data-driven methods are needed to identify key criteria and condense high-dimensional evaluations (Osheyor et al., 2025). Researchers have therefore adopted AI techniques, particularly Machine Learning (ML), to enhance and automate parts of the evaluation. Most recently, advances in generative AI and LLMs have opened the possibility of supporting or even automating expert elicitation within MCDM (Svoboda & Lande, 2024; Wang et al., 2025). To highlight this evolution and position the proposed AIDM framework, the literature is reviewed thematically, moving from data-driven hybrids through criteria-weighting approaches and sustainability extensions to emerging LLM-augmented decision systems.

2.1 Data-Driven Hybrid AI and MCDM Frameworks

Further examples of sustainable and resilient supplier selection methods include (Alavi et al., 2021; Gegovska et al., 2020; Md Muzahid Khan & Bashar, 2023; Renjieh et al., 2010; Nazari-Shirkouhi et al., 2023; Wang et al., 2024). Representative early MCDM works include (Astanti et al., 2020; Dehghanimohammadabadi & Kabadayi, 2020; Kabadayi & Dehghanimohammadabadi, 2022; Li et al., 2020; Roseanna & Saaty, 1987). Foundational hybrid AI/MCDM approaches that couple MCDM with machine learning are illustrated by (Aggarwal et al., 2021; Celebi & Bayraktar, 2008; Golmohammadi et al., 2009; Ha & Krishnan, 2008; Kumar & Roy, 2010; Desheng, 2009). Recent reviews and perspectives on the role of generative AI in decision support and supply-chain management are provided in (Bahrini et al., 2023; Jason et al., 2024; Chuma & Oliveira, 2023; Samuel Fosso Wamba et al., 2024; Tushar Mehrotra Sakshi et al., 2024; Soori et al., 2024; Zuheros et al., 2024).

Hybrid AI/MCDM studies seek to retain the transparency of classical MCDM while leveraging data-driven models to enhance scalability and predictive accuracy. A common strategy is to couple MCDM techniques with supervised ML models that learn supplier performance scores from quantitative and qualitative data. Some hybrid models incorporate dimensionality-reduction techniques to condense numerous evaluation criteria into a smaller set of factors before applying MCDM. For instance, one recent framework applies non-negative matrix factorisation to reduce twenty-four criteria to eight core dimensions and then uses a random forest to derive data-driven weights before ranking suppliers with TOPSIS (Osheyor et al., 2025). Other studies extend hybrid models by combining AHP or Data Envelopment Analysis (DEA) with ANNs, SVMs, and evolutionary algorithms to predict supplier scores and optimise decision parameters. In most cases, the hybrid approach improves ranking accuracy and handles larger datasets, but the set of evaluation criteria and their relative importance remain defined by experts.

A second subset of hybrid research focuses on automated criteria weighting and filtering. Decision-tree classifiers, random forests and other interpretable ML models are used to identify the most influential supplier attributes; AHP or other MCDM methods are then applied to a reduced set of criteria. For instance, Abdulla et al. used a DT to calculate criteria weights and applied AHP using only the most important factors; later work by Abdulla and Baryannis integrated a range of interpretable models (random forests, k-nearest neighbours and extra trees) with AHP to enhance explainability. Other approaches combine fuzzy Delphi methods, ANNs and evolutionary algorithms to elicit weights and evaluate suppliers. These techniques help prioritize criteria and improve transparency, but the initial list of potential attributes is still derived from literature or expert opinion, and the resulting weights vary across decision makers.

Hybrid frameworks have also been extended to address sustainability and resilience. Green supplier selection studies incorporate environmental and social criteria into models such as DEA—analytical network process—ANN hybrids, fuzzy AHP coupled with TOPSIS and Elimination and Choice Translating Reality (ELECTRE), or best—worst method (BWM) combined with fuzzy inference systems. Other work integrates Z-number DEA with ANN to capture uncertainty, or couples BWM with gradient boosting to assess suppliers' resilience and flexibility. These extensions illustrate how hybrid AI/MCDM models can accommodate additional dimensions such as environmental impact and supply-chain risk, but they also exacerbate the complexity of criteria elicitation and amplify the need for reproducible weighting procedures.

Table 1 summarises representative hybrid AI and MCDM supplier selection studies, listing the methods used, the number of suppliers evaluated and the criteria considered. Although

these hybrids vary in their choice of AI techniques, spanning ANNs, SVMs, DTs, logistic regression and meta-heuristics, the common thread is that they augment rather than replace expert judgment. Most depend on decision makers to define criteria and validate results, and few document their elicitation protocols in a way that enables replication.

2.2 Generative AI and LLM-Augmented MCDM Frameworks

Recent advances in generative AI and LLMs have introduced new possibilities for decision support. In this context, multiple studies have begun experimenting with LLMs in multi-criteria settings. For example, Vahidnia (2025) introduced a multi-agent system in which several large language model agents propose criteria weights and negotiate a consensus among themselves, improving weighting efficiency and stability (Mohammad & Vahidnia, 2025). Wang et al. developed a general LLM-based MCDM framework and showed that fine-tuning via Low-Rank Adaptation (LoRA) significantly improves alignment with human expert judgments (Wang et al., 2025). These early results suggest that LLMs can support pairwise comparisons and weighting tasks; however, current experiments often rely on improvised prompting strategies and lack formalised protocols. Beyond supplier selection, LLMs have also been used to create AHP virtual assistants, determine criteria and alternatives, and perform pairwise comparisons.

Despite these promising results, LLM-augmented MCDM remains in its infancy. Most studies rely on ad-hoc prompting strategies and do not provide open, reusable implementations. Furthermore, no published work tailors an LLM-MCDM pipeline specifically to supplier selection or formally evaluates stability across multiple virtual experts. The proposed AIDM framework aims to address these shortcomings by implementing AHP within a documented generative AI pipeline, incorporating an ensemble of virtual expert agents to improve stability, and providing code and configuration artefacts for external verification.

LLM-augmented decision systems and generative AI applications to MCDM and supplier evaluation have been explored in (Zardasht Mahmud Ahamed et al., 2023; Jason et al., 2024; Osheyor et al., 2025; Morales, 2024; Svoboda & Lande, 2024; Wang et al., 2025; Wang & Wu, 2024; Zuheros et al., 2024). In summary, the literature reveals several persistent gaps that motivate this work:

1. **Dependence on expert judgment.** Traditional MCDM methods (e.g., AHP, TOPSIS) rely on human experts to define criteria and weights (Astani et al., 2020; Li et al., 2020; Roseanna & Saaty, 1987). These judgments are subjective and may not scale when many criteria or suppliers are involved.
2. **Limited reproducibility in hybrid AI/MCDM models.** Hybrid frameworks that combine AI with MCDM improve predictive accuracy and handle larger datasets (Abdulla & Baryannis, 2024; Abdulla et al., 2019, 2023; Aggarwal et al., 2021; Celebi & Bayraktar, 2008; Ha & Krishnan, 2008; Kumar & Roy, 2010; Desheng, 2009), yet they still depend on experts for criteria selection and weight assignment. Few studies document their elicitation procedures, making results difficult to replicate across contexts.
3. **Nascent and unstructured use of generative AI.** Early experiments with LLM-augmented MCDM show promise for automating pairwise comparisons and weighting (Zardasht Mahmud Ahamed et al., 2023; Jason et al., 2024; Svoboda & Lande, 2024; Wang et al., 2025; Wang & Wu, 2024; Zuheros et al., 2024), but current research lacks formalised prompting strategies, stability checks and domain-specific guidelines. No existing framework has been tailored to supplier selection or made openly available for replication.

Table 1 Summary of Hybrid AI/MCDM Supplier Selection Studies*

AHP-based Hybrid Models			
Authors	Method	Number of Suppliers	Criteria
Ha and Krishnan	AHP + DEA + ANN	27	production facilities, quality management, claims, delivery, business plans, internal audit
Golmohammadi et al.	AHP + ANN + GA	57	delivery, quality, price, transportation cost, technology, production system
Kumar and Roy	AHP + ANN	7	quality, delay times, unit cost, service
Lakshmanpriya et al.	AHP + ANN	7	quality, cost, competency, capacity, control, cash, consistency, culture, sustainability, communication
Tang et al.	AHP + ANN	4	quality, delay times, unit cost, service
Gegovska et al.	Fuzzy AHP + Fuzzy TOPSIS + Fuzzy ELECTRE + ANN	5	quality, cost, delivery, service, pollution control, green product, environmental management
Abdulla et al.	DT + AHP	unknown	technical acceptance, price, quality
Abdulla and Baryannis	DT/Random Forest/KNN + AHP	3147	quality, price, delivery, payment conditions, offer validity
AI/ML-based Hybrid Models			
Celebi and Bayraktar	DEA + ANN	20	cost, quality, delivery, service
Wu	DEA + DT + ANN	11	quality management, documentation, process capability, management, design, cost reduction
Kuo et al.	DEA + ANP + ANN	12	quality, cost, delivery, environment, service, CSR
Kuo et al.	Fuzzy Delphi + Fuzzy ANN + PSO	17	quality, price, location, finance, delivery, long-term relationship, technical capability, quick response
Fallahpour et al.	DEA + ANN + GP	48	quality, cost, delivery, service, flexibility, customer satisfaction

Table 1 continued

AHP-based Hybrid Models	
Fallahpour et al.	DEA + SVM 48 quality, cost, delivery, service, flexibility, customer satisfaction
Cheng et al.	DEA + TOPSIS + GOA + GP + SVM 26 shipment costs, eco-design, safety, timeliness, billing accuracy
Bahadori et al.	ANN + F-VIKOR - quality, price, delivery, packaging, supplier background, payment terms
Pinar	q-rung fuzzy MCDM + sentiment analysis 4 quality, comfort, ergonomics, price, technical features
Alavi et al.	BWM + FIS 159 cost, quality, delivery, flexibility, technology, environment, social and labor standards
Aggarwal et al.	DEA + DT + SVM + Logistic Regression 500 capacity, finance, management, safety, audit, performance, price, quality
Nazari-Shirkouhi et al.	Z-DEA + ANN 36 delivery, quality, price, technology, risk, responsiveness
Khan et al.	BWM + Gradient Boosting 28 reliability, flexibility, responsiveness, cost, digital technology, services
Wang et al.	AI-powered MCMSS 3 cost, quality, reliability, safety, environment
Abdulla et al.	DT/Random Forest/Extra Trees/CatBoost + MARCOS 4 price, quality, quantity, payment, delivery terms
Subramanian et al.	DT + WENSLO 10 experience, references, technology, capacity, quality, finance, deadlines

GA = Genetic Algorithm; DT = Decision Tree; KNN = k-Nearest Neighbour; ANP = Analytic Network Process; PSO = Particle Swarm Optimization; GP = Genetic Programming; GOA = Grasshopper Optimization Algorithm; FIS = Fuzzy Inference System; CSR = Corporate Social Responsibility; MARCOS = Measurement of Alternatives and Ranking according to Compromise Solution; WENSLO = Weights by Envelope and Slope; MCMSS = Multi-Criteria Material Selection System

These gaps underscore the need for a structured, reproducible approach that leverages generative AI to support expert elicitation while retaining the transparency of classical MCDM. By documenting prompts, aggregating diverse LLM agents into an ensemble of virtual experts, and releasing code and artefacts, the proposed AIDM framework aims to advance beyond existing hybrid AI models and early LLM-MCDM studies.

3 AI-driven Decision-Making Framework

Building on the advancements in AI for decision-making, this paper introduces a new framework that synergizes methodologies from recent studies, specifically those by Wang and Wu and Svoboda and Lande. This proposed AIDM framework is designed to enhance decision accuracy and efficiency across various complex decision scenarios and can be adapted for different applications such as supplier selection, strategic business evaluations, or any other context requiring MCDM. The versatility of this framework allows for the integration of any MCDM model with GenAI technologies, providing a robust tool for complex decision-making processes in a diverse range of industries.

The following subsections delve into the core components of the AIDM framework, detailing the intricacies of its application in simulating and augmenting human expertise in decision-making.

3.1 Expert Identification

As depicted in Figure 1, the AIDM process harnesses a Language Model to virtually replicate the expertise traditionally provided by human experts in the MCDM approach. As part of the AIDM framework, the Expert Identification phase is crucial. This involves:

- **Prompt Development:** Crafting the prompts to establish the decision-making context for the AI.
 - *Context Specification:* Setting up the AI's environment by clearly outlining the problem space.
 - *Persona Crafting:* Creating a detailed virtual expert profile for the AI, enabling it to mimic human expert behavior.
 - *Instruction Configuration:* Providing the AI with precise instructions to guide its decision-making process.
- **Refinement Process:** Enhancing the AI's performance through iterative development.
 - *Testing:* Assessing the AI's decision-making capabilities through simulated scenarios.
 - *Iterative Feedback:* Using the insights gained from testing to refine the AI's decision-making processes.

3.2 Criteria Development

The Criteria Development phase is essential in the AI-Enabled Decision Making framework. It involves:

- **Objective Definition:** Setting clear goals for the AI to address in the decision-making process.

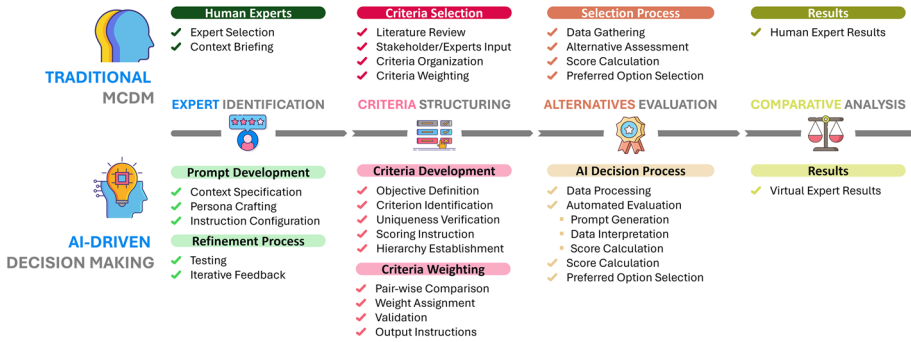


Fig. 1 Detailed comparison of the procedural steps in Traditional MCDM versus the AIDM framework, highlighting the transition from human expert reliance to AI-driven virtual experts in decision support systems

- **Criterion Identification:** Utilizing AI to simulate virtual experts who will identify key criteria.
- **Uniqueness Verification:** Filtering out duplicate criteria to ensure a unique set for evaluation.
- **Scoring Instruction:** Instructing virtual experts to score criteria using a scale where 1 indicates the least and 9 the most importance.
- **Hierarchy Establishment:** Organizing criteria into a hierarchical structure based on their relative significance.

Following the development of criteria, the Criteria Weighting phase is conducted. This phase encompasses:

- **Pair-wise Comparison:** AI, through virtual experts, conducts pair-wise comparisons of criteria based on their importance.
- **Weight Assignment:** The AI system assigns weights to each criterion based on the outcome of the comparisons.
- **Validation:** Validation processes are implemented to ensure the consistency and reliability of the assigned weights.
- **Output Instructions:** Detailed instructions are provided to the AI system to output the findings in a structured format, often an Excel file, for further analysis and verification.

3.3 Alternatives Evaluation

The last subsection is Alternatives Evaluation, which encompasses the AI Decision Process. This phase involves:

- **Data Processing:** Utilize AI to interpret and analyze the information provided in response to the prompts. This involves understanding the context, extracting relevant data, and synthesizing it into a structured format for further analysis.
- **Automated Evaluation:**
 - *Prompt Generation:* Craft prompts to establish the decision-making context for the AI.
 - *Data Interpretation:* Use AI to interpret and analyze the data gathered from the prompts.
 - *Score Calculation:* Calculate scores for each alternative based on the evaluated data.

- **Preferred Option Selection:** Select the preferred option based on the calculated scores and decision criteria.

3.4 Comparative Analysis of Decision-Making Results

In the final phase of the AIDM framework, engagement in a **Comparative Analysis** contrasts the outcomes derived from AI-driven virtual experts with those obtained from human experts. This critical evaluation benchmarks the effectiveness of the AI model and provides insights into the consistency and reliability of the virtual decision-making process, ensuring that the AI's performance aligns with human expert standards.

The integration of ChatGPT and similar GenAI technologies within the AIDM framework marks a significant innovation in the field of decision-making. By harnessing the advanced capabilities of AI, this framework introduces a new paradigm that significantly improves decision accuracy and efficiency. The versatility of AI-driven processes allows for rapid adaptation to diverse decision scenarios, providing a substantial benefit over traditional methods. The iterative and dynamic nature of AIDM also ensures continuous learning and improvement, paving the way for increasingly sophisticated decision support systems.

Moreover, the AIDM framework's flexibility enables it to encompass any MCDM methodology, such as AHP, TOPSIS, Decision Making Trial and Evaluation Laboratory (DEMATEL), among others. GenAI technologies like ChatGPT are adept at emulating these complex decision processes. Advanced versions, such as ChatGPT-4, can analyze original academic or foundational texts, assimilating the nuances of these methodologies to refine the decision-making models. This allows for a nuanced application that aligns closely with the intricate nature of established decision-support frameworks.

The detailed mathematical foundations, input—process—output structure, and system architecture of the AIDM framework are provided in Appendix A.

4 Case Study: AIDM in Supplier Selection

The case study on supplier selection in a heavy machinery manufacturing setting illustrates the AIDM framework's applicability. Recognized within the MCDM domain, supplier selection has been explored using different MCDM approaches, including AHP (Astanti et al., 2020), TOPSIS (Kabadayi & Dehghanimohammadabadi, 2022), and DEMATEL (Li et al., 2020). The AHP method, one of the most applied MCDM models, introduced by Saaty, employs a pairwise comparison method designed to elicit consistent relative judgments (Dehghanimohammadabadi & Kabadayi, 2020).

4.1 Expert Identification

As depicted in the AIDM framework, the first step is to identify experts. The crucial part of this phase is writing a well-defined prompt so that Chat-GPT can provide accurate information based on the problem context. As outlined in Figure 2, which illustrates the prompt development necessary for generating a team of virtual AI agents. These agents, named Expert 1 to Expert 5, are designed to have diverse backgrounds to cover all aspects of supplier selection in the heavy machinery manufacturing sector. The prompt aims to facilitate comprehensive understanding and unique insight from each virtual expert.

Prompt Development

This prompt aims to facilitate the creation of virtual AI agents tasked with streamlining supplier selection processes within the heavy machinery manufacturing industry.

- ✓ **Context Specification:** The AI agents will assist in identifying criteria and suppliers that can provide high-quality and durable <hydraulic systems>. The agents should prioritize suppliers with robust supply chains and a history of timely deliveries.
- ✓ **Persona Crafting:** We aim to create virtual experts embodying the expertise of supply chain managers with extensive knowledge in the <manufacturing sector>. These experts are adept at evaluating suppliers based on some factors that will be defined in the next step.
- ✓ **Instruction Configuration:** Create <five> different agents, each with a slightly different background to offer unique insights during the evaluation process. Name these experts <Expert 1-5>. These experts should cover diverse perspectives. List these experts in a <table format>.

Identified AI Agents by ChatGPT-4

Expert	Background	Area of Expertise	Unique Insight
Expert 1	Extensive experience in hydraulic engineering	Technical assessment of hydraulic systems	Offers in-depth technical analysis of products and materials used
Expert 2	Background in supply chain logistics	Supply chain reliability and efficiency	Focuses on the robustness and resilience of the supplier's supply chain
Expert 3	Worked in quality control for heavy machinery	Quality assurance and control standards	Prioritizes suppliers with the highest quality certifications and reviews
Expert 4	Experience with international suppliers	Global market trends and international sourcing	Brings insights on global supplier networks and risk management
Expert 5	Specialization in sustainable manufacturing	Eco-friendly and sustainable supply practices	Evaluates suppliers based on their sustainability and environmental impact

Refinement Process

In case the expert results are not satisfactory, the prompt can be refined and adjusted based on the feedback received. Modifications will be made iteratively to the prompt to ensure satisfactory outcomes.

Fig. 2 Prompt example for developing virtual AI experts, showcasing the information profile provided by ChatGPT-4. The highlighted parts of the prompt, indicated by < . . . >, are some of the adjustable inputs that can be changed during the refinement process to achieve the desired combination of virtual experts

Criteria Development Prompt

The objective of this prompt is to develop criteria for the supplier selection process using ChatGPT. The criteria will be generated based on the expertise of virtual AI experts, named Expert 1 through Expert 5.

- ✓ **Criterion Identification:** For each of the developed virtual experts (Expert 1, ..., Expert 5), criteria for supplier selection will be developed based on their backgrounds. <8> criteria will be developed for each expert. The results will be saved in an Excel worksheet called "Criterion Identification" with two columns: Expert name and the criteria.
- ✓ **Uniqueness Verification:** Duplicate criteria will be removed, and criteria with similar meanings will be merged. Unique criteria will be listed <max 20>, and the results will be saved into a new tab of Excel called "Unique Criteria".
- ✓ **Scoring Instruction:** Based on the unique criteria results, all virtual experts will rate the importance of each criterion from 1 (least important) to 9 (most important). The results of scoring will be tabulated in a new worksheet, with a column for criteria and a column for each expert (Expert 1, ..., Expert 5). Total scores for each criterion will be calculated as the sum of all scores for that criterion based on experts' ratings. The last column will include the total score for each criterion. Criteria will be sorted based on their total score from largest to smallest and saved in a new worksheet.
- ✓ **Hierarchy Establishment:** The top <15> high scored criteria will be considered and categorized into three categories to create the hierarchy. They will be listed in the Excel file. All worksheets will be part of the same Excel file.

Fig. 3 Criteria development process utilizing ChatGPT-4

4.2 Criteria Development

Once the virtual AI experts are identified, ChatGPT can be applied to provide a list of important criteria for the selected problem, namely supplier selection. Leveraging the capabilities of ChatGPT-4 and its API, all of the Criteria Development tasks are completed within a single prompt. The prompt comprehensively guides the AI through the process, ensuring that all relevant aspects are covered. As depicted in Figure 3, this prompt facilitates the creation of a structured list of criteria that are essential for evaluating potential suppliers.

The hierarchy of supplier selection criteria is a crucial aspect of the decision-making process. As illustrated in Figure 4, the criteria have been categorized into three primary dimensions: Quality, Cost, and Attributes. Each of these dimensions is further broken down into specific factors that contribute to the overall assessment of potential suppliers. The Quality dimension encompasses criteria such as the inherent quality of goods or services, reliability, communication, production flexibility, and customer feedback. The Cost dimen-

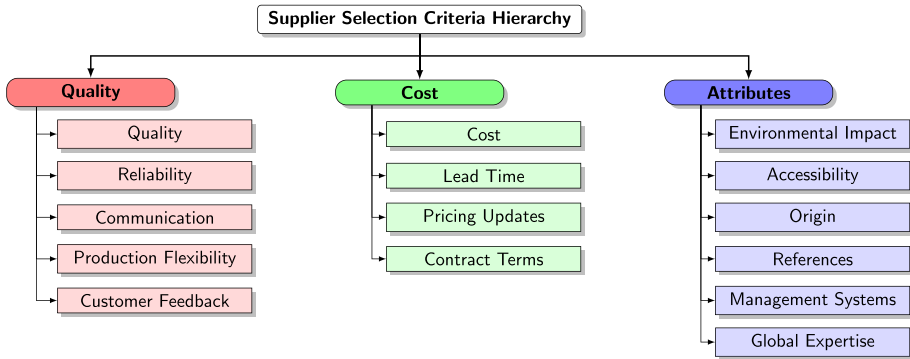


Fig. 4 Supplier Selection Criteria Hierarchy

sion includes direct cost, lead time, frequency of pricing updates, and contract terms. Lastly, the Attributes dimension captures environmental impact, accessibility, origin of the supplies, references provided, management systems in place, and the global expertise of the supplier. This structured approach aids in systematically evaluating and comparing suppliers to select the most suitable one based on a comprehensive set of criteria.

The subsequent step is to perform the pairwise comparison and assign weights, a meticulous task when carried out manually, as it requires each expert to evaluate and score criteria against each other. Additionally, considerations such as the Consistency Ratio (CR) must be taken into account to ensure the validity of the assessments.

With the advent of Generative AI, this process can be significantly optimized. As illustrated in Figure 5, a well-crafted prompt with ChatGPT-4 can perform all tasks associated with AHP, including conducting the analysis and performing the calculations necessary to obtain the weights. Notably, since ChatGPT-4 is employed, the platform has access to a wealth of online resources and can understand the procedure of the calculation. It can process inputs such as a diagram of criteria hierarchy or the table of expert evaluations to provide weights corresponding to the defined problem.

This approach does not only complete the step-by-step calculation but also generates a detailed Excel file that includes all the details required for verification and validation. Should the results not meet the expected standards, the prompt can be adjusted until satisfactory results are achieved. This makes the calculation process not only efficient but also adaptable, ensuring the methodology aligns with the nuanced requirements of the decision-making scenario. The refinement process follows iterative prompt adjustment based on output quality metrics including consistency ratio compliance, domain relevance assessment, and expert profile diversity validation.

As part of the Criteria Weighting phase, a snapshot of the criteria pairwise comparison provided by the Generative Pre-trained Transformer 4 is depicted in Figure 6. The table showcases the varied perspectives of different virtual experts on the importance of each criterion relative to others within the 'Quality' dimension. This exemplifies the use of the platform to simulate a comprehensive expert panel evaluation, streamlining the Analytic Hierarchy Process.

Criteria Weighting Prompt

Conduct a revised Analytic Hierarchy Process (AHP) developed by Saaty in 1980 for supplier selection based on updated criteria and expert inputs. The goal is to assess suppliers for heavy machinery manufacturing by comparing multiple criteria across quality, cost, and other attributes. Assume that I will provide the list of criteria and the background information for each expert.

- ✓ **Pair-wise Comparison:** Each expert will compare the criteria within their specific dimension of expertise using Saaty's scale. The comparisons will be between pairs of criteria related to Quality, Cost, and other Attributes. I will provide you with the list of criteria and the experts' specific areas of expertise.
- ✓ **Weight Assignment:** Assign weights to the pairs of criteria based on the experts' comparisons. Utilize Saaty's scale where 1 represents equal importance and 9 represents extreme importance, with intermediate values for nuanced preferences. The weights should reflect the experts' assessments based on their backgrounds.
- ✓ **Validation:** Ensure the consistency of the pairwise comparisons and weights. Validate that the weights are reasonable and align with the expertise of the relevant expert. If inconsistencies are detected, indicate the need for reassessment.
- ✓ **Output Instructions:** Generate an Excel worksheet that includes the following sheets:
 - 1) Pairwise Comparisons: For each expert's scores.
 - 2) Aggregated Comparison Matrix: Combining scores from all experts.
 - 3) Normalized Comparison Matrix: To obtain comparable weights.
 - 4) Criteria Weights: Calculating the final weights of each criterion.
 - 5) Provide the file in a downloadable format and advise on any inconsistencies found during validation.

Fig. 5 The Analytic Hierarchy Process facilitated by ChatGPT-4, showcasing the streamlined calculation and validation process

	A	B	C	D	E	F	G	H	I
1	Dimension	Criterion 1	Criterion 2	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Geometric Average
2	Quality	Quality	Reliability	1.40	1.00	1.29	1.00	1.00	1.12
3	Quality	Quality	Communication	7.00	0.33	9.00	1.00	1.00	1.84
4	Quality	Quality	Production Flexibility	7.00	1.00	9.00	1.00	1.00	2.29
5	Quality	Quality	Customer Feedback	7.00	1.00	3.00	1.00	1.00	1.84
6	Quality	Quality	Cost	2.33	0.14	9.00	1.00	1.00	1.25
7	Quality	Quality	Lead Time	7.00	0.20	9.00	1.00	1.00	1.66

Fig. 6 A snapshot of the pairwise comparison for 'Quality' criteria as generated by ChatGPT

4.3 Evaluation

Once the criteria are determined and weights assigned, alternative evaluation can be performed. Regardless of the approach, whether through human expertise or virtual AI experts, the evaluation involves a systematic comparison of potential suppliers against the established criteria. This section, however, is omitted from this article since the primary aim of the paper is to demonstrate the application of AIDM for determining criteria and weights. The selection process itself can be conducted using any specific MCDM approach. The focus here is to highlight the efficiency and effectiveness of AIDM in the critical preparatory steps that precede the actual supplier selection, ensuring that the subsequent MCDM process is grounded on a robust and well-defined set of criteria.

4.4 Statistical Comparative Analysis

This section compares AIDM and human expert criterion weights using absolute and relative differences across domestic, international, and overall groups, with inference based on non-parametric Wilcoxon signed-rank tests. To ensure accurate comparisons, criteria were first semantically aligned using Jaccard similarity (a pair was accepted if it shared more than two content words).

The human expert panel comprised 14 purchasing department professionals from a heavy machinery manufacturing organization, divided into domestic (n=8) and international (n=6) purchasing teams. While detailed individual profiles were not available due to confidentiality constraints, the experts represented operational decision-makers with direct supplier selection responsibilities within their respective geographical domains.

4.4.1 Consistency Ratio Analysis

The consistency ratio analysis in Table 2 evaluates the logical coherence of pairwise comparisons across both human experts and AIDM virtual experts. Both methodologies demonstrated acceptable consistency ratios well below the 0.10 threshold, though human experts achieved superior consistency performance across all comparison levels.

The AIDM framework's re-iteration protocol successfully maintained consistency quality, with 85% of virtual experts achieving acceptable CR values. The slightly higher AIDM consistency ratios reflect the inherent challenge of maintaining logical coherence across AI-generated pairwise comparisons, though all values remain within acceptable professional practice standards. The domestic human experts demonstrated exceptional consistency (CR = 0.030), suggesting high agreement and logical coherence in their pairwise evaluations. International human experts showed moderate consistency (CR = 0.060), which may reflect greater diversity in perspectives or decision-making approaches across different cultural and operational contexts.

4.4.2 Domestic Expert Comparison

In the domestic set, systematic differences were observed between human purchasing professionals and AIDM virtual experts across the twelve aligned criteria (Table 3). Human judgments concentrated weight on execution-oriented factors, with notably higher allocations to *Lead Time* (0.123 vs. 0.054), *Maintaining Price Updates* (0.099 vs. 0.031), *Reliability* (0.114 vs. 0.064), and *Signing Contracts* (0.072 vs. 0.030), reflecting a focus on timely delivery, cost control, and managing contract risks in routine purchasing. By contrast, AIDM assigned greater weight to strategic and compliance-adjacent attributes, including *Origin Country* (0.065 vs. 0.029), *References* (0.054 vs. 0.025), *Accessibility* (0.071 vs. 0.046), *Environmental Awareness* (0.050 vs. 0.024), and *Global Company Skills* (0.038 vs. 0.021), indicating broader consideration of network risk, sustainability, and organizational scalability. Differences in *Management Systems Documents* were small (0.042 vs. 0.036), and *Production Opportunities* showed only modest divergence (0.032 vs. 0.045). Overall, domestic human experts emphasized a narrower set of operational levers, whereas AIDM produced a more balanced profile that elevates sustainability issues. These complementary emphases suggest that, in practice, human panels can focus on immediate practicality, while AIDM highlights early signs of country risk, regulatory issues, and supplier capacity that are often overlooked in routine reviews.

4.4.3 International Expert Comparison

In the international set, patterns are similar to the domestic results but shift due to the added challenges of international sourcing (Table 4). Human experts assign the highest weight to *Reliability* (0.139 vs. 0.075) and continue to emphasize *Maintaining Price Updates* (0.107 vs. 0.033), *Accessibility* (0.094 vs. 0.073), *Communication* (0.093 vs. 0.041), and *Signing Contracts* (0.087 vs. 0.030), indicating a focus on timely delivery, cost control, and managing contract risks in routine purchasing across dispersed supplier networks. By contrast, AIDM allocates more weight to *Origin Country* (0.066 vs. 0.024), *Environmental Awareness* (0.051 vs. 0.016), and *Management Systems Documents* (0.053 vs. 0.027), elevating country risk, sustainability, and documentation quality. *Lead Time* is nearly aligned (0.056 vs. 0.061), and *Production Opportunities* is identical (0.033 vs. 0.033), suggesting convergence on core

Table 2 Consistency Ratio comparison between human experts and AIDM

Analysis Level	Human Expert CR	AIDM Average CR	AIDM Std. Dev.	CR Assessment
Domestic	0.030	0.082	0.041	Both Acceptable
International	0.060	0.074	0.028	Both Acceptable
Overall	0.045	0.078	0.034	Both Acceptable

Note: Acceptable CR threshold ≤ 0.10

Table 3 Domestic expert vs AIDM criteria weight comparison.

Criterion	Human Weight	AIDM Weight	Absolute Diff.	Relative Diff.
Accessibility of the supplier	0.046	0.071	0.025	0.543
Reliability of the supplier	0.114	0.064	0.050	0.439
Lead Time	0.123	0.054	0.069	0.561
Origin country of the supplier	0.029	0.065	0.036	1.241
References of the supplier	0.025	0.054	0.029	1.160
Management systems documents	0.036	0.042	0.006	0.167
Communication skills	0.054	0.033	0.021	0.389
Environmental awareness	0.024	0.050	0.026	1.083
Global Company Skills	0.021	0.038	0.017	0.810
Production Opportunities	0.045	0.032	0.013	0.289
Maintaining Price Updates	0.099	0.031	0.068	0.686
Signing contracts	0.072	0.030	0.042	0.585

Table 4 International expert vs AIDM criteria weight comparison.

Criterion	Human Weight	AIDM Weight	Absolute Diff.	Relative Diff.
Accessibility of the supplier	0.094	0.073	0.021	0.223
Reliability of the supplier	0.139	0.075	0.064	0.460
Lead Time	0.061	0.056	0.005	0.082
Origin country of the supplier	0.024	0.066	0.042	1.750
References of the supplier	0.050	0.056	0.006	0.120
Management systems documents	0.027	0.053	0.026	0.963
Communication skills	0.093	0.041	0.052	0.559
Environmental awareness	0.016	0.051	0.035	2.188
Global Company Skills	0.030	0.040	0.010	0.333
Production Opportunities	0.033	0.033	0.000	0.000
Maintaining Price Updates	0.107	0.033	0.075	0.695
Signing contracts	0.087	0.030	0.057	0.654

Table 5 Wilcoxon signed-rank test results for statistical significance assessment.

Analysis Level	Sample Size	Wilcoxon Statistic	p-value	Statistical Significance
Domestic	12	29	0.470	Not Significant
International	12	25	0.301	Not Significant
Overall	12	26	0.339	Not Significant

Note: Statistical significance assessed at $\alpha = 0.05$ level

capacity signals even as priority ordering differs elsewhere. Taken together, international human judgments concentrate on immediate practicality under global uncertainty, while AIDM highlights early signs of country risk, regulatory issues, and supplier capacity that may be overlooked in routine reviews.

4.4.4 Statistical Significance Testing

Wilcoxon signed-rank tests were applied to paired criterion weights (AIDM vs. human) to assess whether observed differences reflect systematic shifts or random variation. Across the domestic, international, and overall comparisons, no test reached the $\alpha = 0.05$ threshold (Table 5). This outcome indicates broad comparability in the distributions of weights across approaches in this sample. Accordingly, interpretation in this section emphasizes the directional patterns documented above (e.g., higher human emphasis on *Reliability* and *Maintaining Price Updates*, and higher AIDM emphasis on *Origin Country* and *Environmental Awareness*) rather than formal statistical rejections.

5 Discussion

The consistency ratio and weighting analyses indicate that AIDM functions as a feasible decision-support approach for supplier selection. Acceptable CR values were observed for both human experts and AIDM, suggesting comparable logical coherence of pairwise judg-

ments. Across domestic, international, and overall comparisons, Wilcoxon signed-rank tests did not reach the $\alpha = 0.05$ threshold, indicating no statistically significant differences in the distributions of weights; however, systematic criterion-level shifts were evident; human panels placed greater emphasis on *Reliability*, *Maintaining Price Updates*, *Communication*, and *Signing Contracts*, whereas AIDM assigned relatively higher weight to *Origin Country*, *Environmental Awareness*, *Management Systems Documents*, and *Accessibility*. These shifts can yield different supplier rankings: human profiles tend to favor cost-competitive, dependable suppliers, while the AIDM profile can elevate suppliers with stronger provenance, sustainability practices, documentation quality, and network reach even when associated with greater complexity. Accordingly, alignment with organizational priorities is advisable, cost-focused manufacturers may lean toward the human profile, whereas sustainability-oriented or globally expanding firms may benefit from the AIDM profile. The absence of significant distributional differences also supports a hybrid implementation in which human operational expertise is complemented by AIDM's strategic signals to balance immediate practicality with longer-term objectives.

Large language models are susceptible to hallucination phenomena, where AI systems generate plausible but incorrect or non-existent information with apparent confidence. In the context of supplier selection criteria, this risk manifests through the potential creation of fabricated industry standards, non-existent certification requirements, or artificially constructed performance benchmarks that appear credible but lack empirical foundation. Mitigation strategies implemented include integration of human-defined criteria as validation approach, systematic consistency checking across AI-generated outputs, and transparent documentation of prompt engineering and generation processes. Despite these controls, the systematic differences between human and AIDM approaches require careful interpretation considering potential AI bias factors that may influence practical implementation decisions.

The current analysis provides essential foundation data for future AIDM refinement and calibration efforts. Enhanced training protocols could achieve closer alignment with human expert preferences while maintaining acceptable consistency standards and preserving the framework's capacity for systematic criteria generation and methodological innovation. Moreover, criteria completeness cannot be guaranteed through the current methodology. The framework relies on expert diversity and literature cross-validation to achieve reasonable coverage while acknowledging potential gaps in domain-specific factors. Future development should explore hybrid approaches that combine AIDM's comprehensive methodology with human expert practical knowledge through iterative calibration processes, potentially enabling decision support systems that leverage both AI capability and human expertise for enhanced supplier selection outcomes.

6 Conclusion

This paper introduced and demonstrated the AIDM framework through a supplier selection case study. By integrating generative AI with the Analytic Hierarchy Process (AHP), the framework showed that large language models can contribute structured judgments comparable to those of human experts.

The contributions are threefold. First, a formalized GenAI—MCDM pipeline was articulated that operationalizes AHP with defined procedures and documented prompting. Second, practical applicability was demonstrated in supplier selection, a core procurement decision.

Third, an open and reusable implementation was provided to enable replication and extension in other contexts.

The empirical analyses indicated acceptable consistency ratios for both human panels and AIDM, and Wilcoxon signed-rank tests did not reach the $\alpha = 0.05$ threshold across domestic, international, and overall comparisons, suggesting broad comparability in the distributions of weights. At the same time, interpretable directional differences were observed (e.g., higher human emphasis on reliability, price updates, communication, and contracting; higher AIDM emphasis on origin country, environmental awareness, documentation, and accessibility).

Overall, AIDM represents a step toward augmenting human judgment with generative AI in complex decision-making settings. By emphasizing methodological rigor, transparency, and reusability, this study contributes to research on AI-supported decision systems and offers actionable guidance for scholars and practitioners.

Funding Open access funding provided by Northeastern University Library

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A Reproducibility

This appendix documents the exact workflow and interface screens used to reproduce the reported results. As shown in Figure 7, the AIDM process executes six modules in sequence from input definition to report export. The open-source implementation (Python, Gradio UI, AHP eigenvalue/CR routines, and Excel exporter) is available [Here](#).

Step 1 — Expert Generation

The workflow begins by specifying the decision topic, domain, number of virtual experts (e.g., 3–5 recommended), and any optional context. The system then generates diverse expert profiles (background, role, geography, education) to cover the decision space. These agents are used downstream for criteria elicitation and AHP judgments (Figure 8).

Step 2 — Criteria Collection

Each expert proposes 5–15 criteria aligned with its specialization. The system consolidates, semantically de-duplicates, and attributes criteria to their sources, forming a unique set for analysis. Typical parameters include “criteria per expert” (e.g., 8) and optional important-factor guidance (Figure 9).

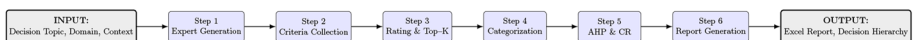


Fig. 7 AIDM workflow

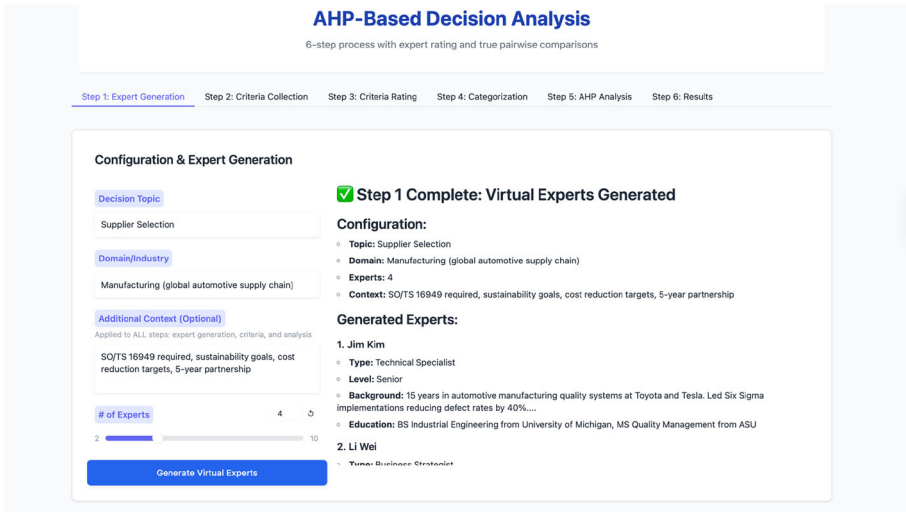


Fig. 8 Step 1-Expert Generation

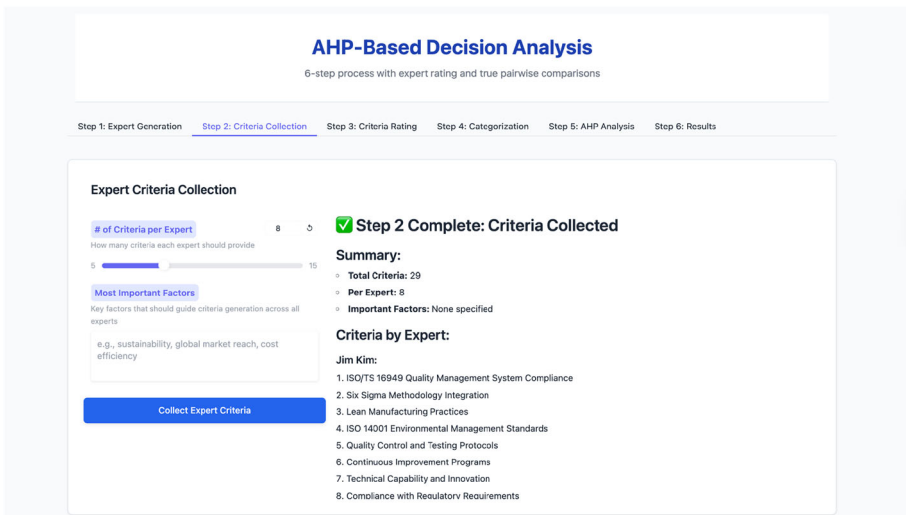


Fig. 9 Step 2-Criteria Collection

Step 3 — Rating & Top-K

Experts rate all criteria on a 1–10 importance scale; the system aggregates ratings and retains the Top-K (e.g., 10–15) for downstream hierarchy and AHP. This step improves tractability while preserving consensus signal strength. Key parameters include Top-K count and rating aggregation (Figure 10).

Step 4 — Categorization

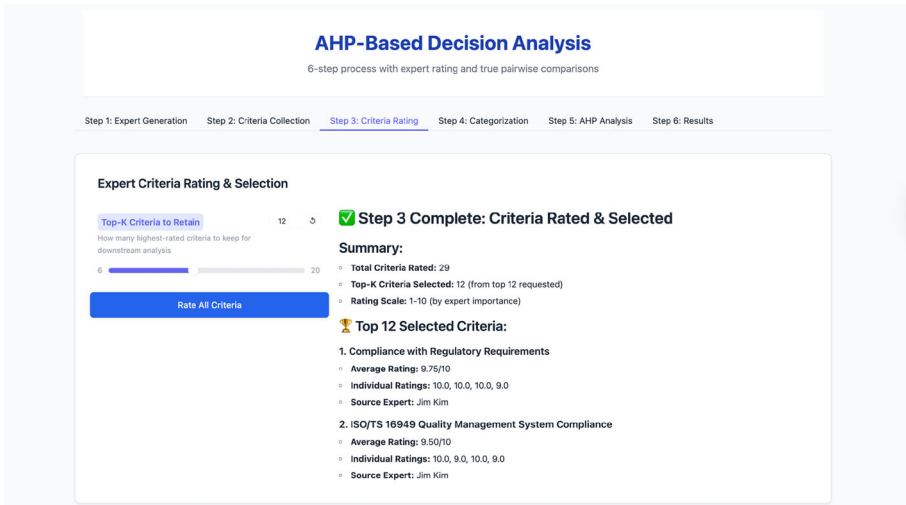


Fig. 10 Step 3-Rating and Top–K Selection

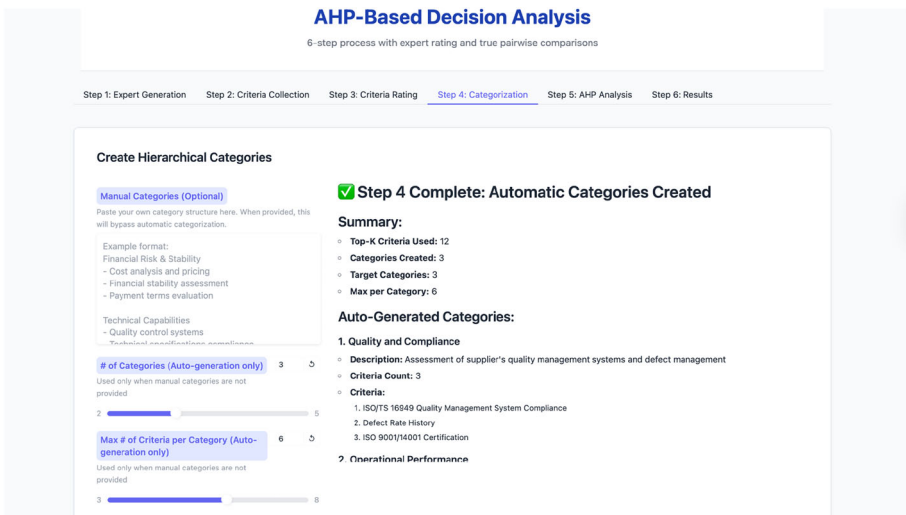


Fig. 11 Step 4-Categorization into logical groups

The selected criteria are grouped into 2–5 categories to form the analysis hierarchy. The tool supports automatic grouping (semantic clustering) and optional manual taxonomy overrides. Typical parameters include category count (e.g., 3–4) (Figure 11).

Step 5 — AHP & Consistency Ratio (CR)

Pairwise comparisons are executed at the category and within-category levels using the eigenvalue method. The consistency ratio (CR) is validated against a user-defined threshold (e.g., 0.10), with automatic re-iterations (e.g., up to 5) when needed. Outputs include local/global weights and CR diagnostics per matrix (Figure 12).

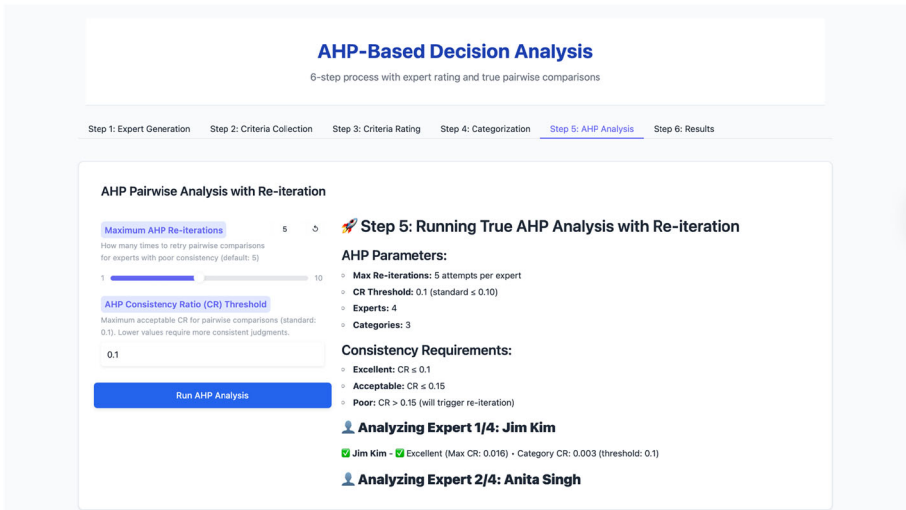


Fig. 12 Step 5-AHP pairwise comparisons with CR checks

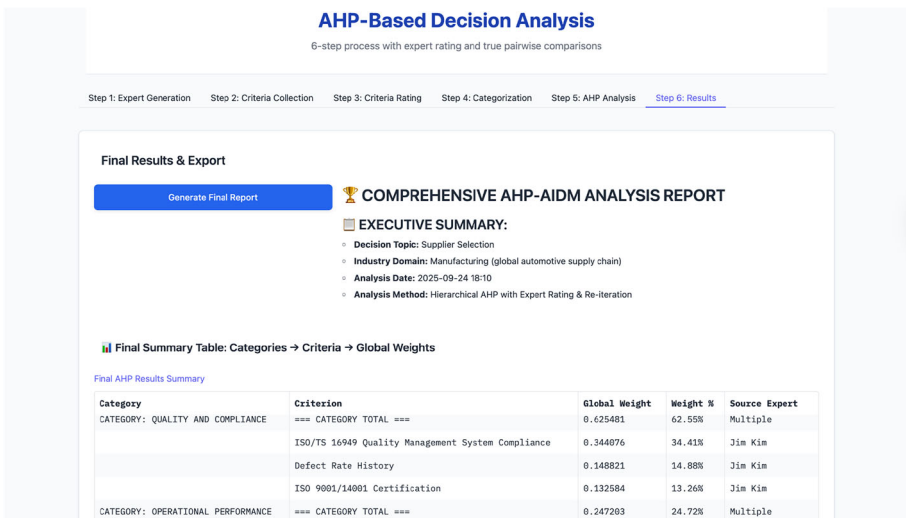


Fig. 13 Step 6-Results and Excel export

Step 6 — Report Generation

The final screen presents consolidated results, including local/global weights and consistency metrics. The Excel exporter generates a multi-worksheet report (e.g., 12 tabs) with a complete audit trail suitable for verification and reuse in other studies (Figure 13).

References

- Abdulla, A., & Baryannis, G. (2024). A Hybrid Multi-Criteria Decision-Making and Machine Learning Approach for Explainable Supplier Selection. *Supply Chain Analytics*, 7, Article 100074.
- Abdulla, A., Baryannis, G., & Badi, I. (2019). *Weighting the Key Features Affecting Supplier Selection Using Machine Learning Techniques*.
- Abdulla, A., Baryannis, G., & Badi, I. (2023). An Integrated Machine Learning and MARCOS Method for Supplier Evaluation and Selection. *Decision Analytics Journal*, 9, Article 100342.
- Aggarwal, I., Gunreddy, N., John, A., & Rajan. (2021). A Hybrid Supplier Selection Approach Using Machine Learning and Data Envelopment Analysis. In *2021 Innovations in Power and Advanced Computing Technologies (I-PACT)*, pages 1–5. IEEE.
- Ahamed, Z. M., Dhahir, H. M., Mohammed, M. M., Ali, R. H., Hassan, S. H., Muhiyaldeen, A. S., Saeed, Y. A., Fatah, M. L., Qaradakh, A. J., Ali, R. M., et al. (2023). Comparative Analysis of ChatGPT and Human Decision-Making in Thyroid and Neck Swellings: A Case-Based Study. *Barw Medical Journal*.
- Alavi, B., Tavana, M., & Mina, H. (2021). A Dynamic Decision Support System for Sustainable Supplier Selection in Circular Economy. *Sustainable Production and Consumption*, 27, 905–920.
- Astanti, R., Mbolla, S., & Ai, T. (2020). Raw Material Supplier Selection in a Glove Manufacturing: Application of AHP and Fuzzy AHP. *Decision Science Letters*, 9(3), 291–312.
- Bahadori, M., Hosseini, S. M., Teymourzadeh, E., Ravangard, R., Raadabadi, M., & Alimohammadzadeh, K. (2020). A Supplier Selection Model for Hospitals Using a Combination of Artificial Neural Network and Fuzzy VIKOR. *International Journal of Healthcare Management*, 13(4), 286–294.
- Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R. J., Esmaceli, M., Majdabadkohne, R. M., & Pashvar, M. (2023). ChatGPT: Applications, Opportunities, and Threats. In *2023 Systems and Information Engineering Design Symposium (SIEDS)*, pages 274–279. IEEE.
- Jason W Burton, Ezequiel Lopez-Lopez, Shahar Hechtlinger, Zoe Rahwan, Samuel Aeschbach, Michiel A Bakker, Joshua A Becker, Aleks Berditchevskaia, Julian Berger, Levin Brinkmann, et al. How Large Language Models Can Reshape Collective Intelligence. *Nature Human Behaviour*, pages 1–13, 2024.
- Celebi, D., & Bayraktar, D. (2008). An Integrated Neural Network and Data Envelopment Analysis for Supplier Evaluation under Incomplete Information. *Expert Systems with Applications*, 35(4), 1698–1710.
- Cheng, Y., Peng, J., Xin, G., Zhang, X., Liu, W., Zhou, Z., Yang, Y., & Huang, Z. (2020). An Intelligent Supplier Evaluation Model Based on Data-Driven Support Vector Regression in Global Supply Chain. *Computers & Industrial Engineering*, 139, Article 105834.
- Chuma, E. L., & De Oliveira, G. G. (2023). Generative AI for Business Decision-Making: A Case of ChatGPT. *Management Science and Business Decisions*, 3(1), 5–11.
- Dehghanimohammadabadi, M., & Kabadayi, N. (2020). A Two-Stage AHP Multi-Objective Simulation Optimization Approach in Healthcare. *International Journal of the Analytic Hierarchy Process*, 12(1).
- Alireza Fallahpour, Ezutah Udoney Ologu, Siti Nurmayana Musa, Dariush Khezrimotlagh, and Kuan Yew Wong. An Integrated Model for Green Supplier Selection under Fuzzy Environment: Application of Data Envelopment Analysis and Genetic Programming Approach. *Neural Computing and Applications*, 27:707–725, 2016.
- Fallahpour, A., Kazemi, N., Molani, M., Nayyeri, S., & Ehsani, M. (2018). An Intelligence-Based Model for Supplier Selection Integrating Data Envelopment Analysis and Support Vector Machine. *Interdisciplinary Journal of Management Studies*, 11(2), 209–241.
- Wamba, S. F., Guthrie, C., Queiroz, M. M., & Minner, S. (2024). ChatGPT and Generative Artificial Intelligence: An Exploratory Study of Key Benefits and Challenges in Operations and Supply Chain Management. *International Journal of Production Research*, 62(16), 5676–5696.
- Gegovska, T., Koker, R., & Cakar, T. (2020). Green Supplier Selection Using Fuzzy Multiple-Criteria Decision-Making Methods and Artificial Neural Networks. *Computational Intelligence and Neuroscience*, 2020(1), 8811834.
- Gidiagba, O. J., Tartibu, L., & Okwu, M. (2025). Integrating machine learning with multi-criteria decision-making models for sustainable supplier selection in dynamic supply chains. *Logistics*, 9(4), 152.
- Golmohammadi, D., Creese, R. C., Valian, H., & Kolassa, J. (2009). Supplier Selection Based on a Neural Network Model Using Genetic Algorithm. *IEEE Transactions on Neural Networks*, 20(9), 1504–1519.
- Ha, S. H., & Krishnan, R. (2008). A Hybrid Approach to Supplier Selection for the Maintenance of a Competitive Supply Chain. *Expert Systems with Applications*, 34(2), 1303–1311.
- Kabadayi, N., & Dehghanimohammadabadi, M. (2022). Multi-Objective Supplier Selection Process: A Simulation-Optimization Framework Integrated with MCDM. *Annals of Operations Research*, pages 1–23.

- Khan, M. M., Bashar, I., Minhaj, G. M., Wasi, A. I., & Hossain, N. U. I. (2023). Resilient and Sustainable Supplier Selection: An Integration of SCOR and Machine Learning Approach. *Sustainable and Resilient Infrastructure*, 8(5), 453–469.
- Kumar, J., & Roy, N. (2010). A Hybrid Method for Vendor Selection Using Neural Network. *International Journal of Computer Applications*, 11(12), 35–40.
- Kuo, R. J., Wang, Y. C., & Tien, F. C. (2010). Integration of Artificial Neural Network and MADA Methods for Green Supplier Selection. *Journal of Cleaner Production*, 18(12), 1161–1170.
- Kuo, R. J., Hong, S. Y., & YC1201 Huang. (2010). Integration of Particle Swarm Optimization-Based Fuzzy Neural Network and Artificial Neural Network for Supplier Selection. *Applied Mathematical Modelling*, 34(12), 3976–3990.
- Lakshmanpriya, C., Sangeetha, N., & Lavanpriya, C. (2013). Vendor Selection in Manufacturing Industry Using AHP and ANN. *The SIJ Transactions on Industrial, Financial & Business Management*, 1(1), 29–34.
- Li, Y., Diabat, A., & Chung-Cheng, L. (2020). Leagile Supplier Selection in Chinese Textile Industries: A DEMATEL Approach. *Annals of Operations Research*, 287(1), 303–322.
- Morales, A. A. S. (2024). Gen-AI and the Future of Supply Chain Management: The Impact of Large Language Models on Modern Supply Chain Management. Master's thesis, Universidade NOVA de Lisboa (Portugal).
- Nazari-Shirkouhi, S., Tavakoli, M., Govindan, K., & Mousakhani, S. (2023). A Hybrid Approach Using Z-Number DEA Model and Artificial Neural Network for Resilient Supplier Selection. *Expert Systems with Applications*, 222, Article 119746.
- Pandey, M., Litoriya, R., & Pandey, P. (2024). Indicators of AI in Automation: An Evaluation Using Intuitionistic Fuzzy DEMATEL Method with Special Reference to ChatGPT. *Wireless Personal Communications*, pages 1–21.
- Pinar, A. (2025). An Integrated Sentiment Analysis and Q-Rung Orthopair Fuzzy MCDM Model for Supplier Selection in E-Commerce: A Comprehensive Approach. *Electronic Commerce Research*, 25(2), 1311–1342.
- Saaty, R. W. (1987). The Analytic Hierarchy Process-What It Is and How It Is Used. *Mathematical Modelling*, 9(3–5), 161–176.
- Sakshi, T. M., Tyagi, P., & Jain, V. (2024). Emerging Trends in Hybrid Information Systems Modeling in Artificial Intelligence. *Hybrid Information Systems: Non-Linear Optimization Strategies with Artificial Intelligence*, page 115.
- Soori, M., Jough, F. K. G., Dastres, R., & Arezoo, B. (2024). AI-Based Decision Support Systems in Industry A Review. *Journal of Economy and Technology*.
- Subramanian, S. P., Pandian, P., Sivaprakasam, R., & Kadarkarai, J. (2025). A Novel Machine Learning Framework for Optimized Supplier Selection Using the Weights by ENvelope and SLOpe (WENSLO) Technique. *Applied Operations and Analytics*, 1(1), 1–10.
- Svoboda, I., & Lande, D. (2024). Enhancing Multi-Criteria Decision Analysis, with AI: Integrating Analytic Hierarchy Process and GPT-4 for Automated Decision Support. *arXiv preprint arXiv:2402.07404*.
- Tang, S. H., Hakim, N., Khaksar, W., Ariffin, M. K. A., Sulaiman, S., & Pah, P. S. (2013). A Hybrid Method Using Analytic Hierarchical Process and Artificial Neural Network for Supplier Selection. *International Journal of Innovation, Management and Technology*, 4(1), 109–111.
- Vahidnia, M. H. (2025). Multi-agent systems of large language models as weight assigners: An approach to collaborative weighting in spatial multi-criteria decision-making. *Geomatica*, page Article 100071.
- Wang, H., Zhang, F., & Chaoxu, M. (2025). *One for All: A General Framework of LLMs-Based Multi-Criteria Decision Making on Human Expert Level.*, arXiv preprint [arXiv:2502.15778](https://arxiv.org/abs/2502.15778).
- Wang, X., & Xiaojun, W. (2024). Can ChatGPT Serve as a Multi-Criteria Decision Maker?. A Novel Approach to Supplier Evaluation. *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 10281–10285. IEEE.
- Wang, Z.-J., Chen, Z.-S., Qin, S., Chin, K.-S., Pedrycz, W., & Skibniewski, M. J. (2024). Enhancing the Sustainability and Robustness of Critical Material Supply in Electrical Vehicle Market: An AI-Powered Supplier Selection Approach. *Annals of Operations Research*, 342(1), 921–958.
- Desheng, W. (2009). Supplier Selection: A Hybrid Model Using DEA, Decision Tree and Neural Network. *Expert Systems with Applications*, 36(5), 9105–9112.
- Zuheros, C., Herrera-Poyatos, D., Montes, R., & Herrera, F. (2024). Large Language Models for Crowd Decision Making Based on Prompt Design Strategies Using ChatGPT: Models. *Analysis and Challenges.*, arXiv preprint [arXiv:2403.15587](https://arxiv.org/abs/2403.15587).