

A Multi-Criteria Decision Support Framework for Artificial Intelligence Project and Large Language Model Selection

Gül Mine Ak

MSc Student, Industrial Engineering Department
Galatasaray University, Istanbul, Türkiye
gul.ak@ogr.gsu.edu.tr

Gülçin Büyüközkan

Professor, Industrial Engineering Department
Galatasaray University, Istanbul, Türkiye
gbuyukozkan@gsu.edu.tr, gulcin.buyukozkan@gmail.com

Abstract

The rapid development of artificial intelligence (AI) technologies, particularly Large Language Models (LLMs), in recent years has radically transformed organizational processes; however, this development has also created a complex management problem for decision-makers regarding how to identify the right project topic and select the most suitable model. The aim of this study is to provide an analytical decision support framework for identifying an AI project topic aligned with corporate strategies and selecting the most suitable LLM for that project. Within the scope of the study, an integrated hierarchical structure encompassing technical, financial, strategic, and ethical dimensions was developed by integrating current studies in the literature and expert opinions. The Analytic Hierarchy Process (AHP), a Multi-Criteria Decision Making method that enables the consistent quantification of multi-dimensional judgments in complex decision settings, was used to solve the decision problem. The analysis of the data obtained from five expert participants revealed that the “Security and Ethics” criterion had a much higher weight than the technical performance and cost criteria in both project selection (47%) and model selection (50%). When the alternatives were evaluated in the empirical application, the “Media and Content Management” project was determined as the most suitable investment due to its strategic alignment, and the “Llama (Meta)” model was selected as the most suitable technical infrastructure due to its data sovereignty and on-premises deployment capability. This study contributes to the literature by adding case-based evidence to an emerging consensus that AI investments are increasingly guided by a security- and ethics-oriented approach rather than by a cost-focused approach. Its main differentiator is the coupling of AI project selection and LLM selection decisions within a single AHP hierarchy.

Keywords

Large Language Models (LLMs); Analytic Hierarchy Process; Multi-Criteria Decision-Making; AI Project Selection; LLM Selection

1. Introduction

Artificial intelligence (AI) technologies, and especially Large Language Models (LLMs), have created a profound transformation across a wide range of fields, from academic research to industrial applications, with their exponential growth in recent years (Li 2025). According to Web of Science data (topic search “LLM” OR “large language models,” retrieved November, 2025) the number of studies focused solely on LLMs exceeds 38,000, which is a concrete indicator of the strategic importance of this technology. However, these broad possibilities offered by the technology

have also brought with them a complex management problem for decisionmakers and organizations in terms of "identifying the right project topic" and "selecting the most appropriate model."

For organizations, artificial intelligence projects are multi-dimensional investments involving high costs, technical uncertainties, data security risks, and ethical concerns. This situation, which can be characterised as a 'selection paradox,' necessitates that decision-makers focus not only on technical performance (speed, accuracy) but also on conflicting criteria such as cost, legal compliance, and societal impact — a multi-dimensionality documented across both classical project selection (Souza et al. 2021; Kandakoglu et al. 2023) and emerging LLM evaluation literature (Şimşek 2025). Traditional and intuitive decision-making methods are insufficient to manage the next-generation risks associated with LLMs, such as hallucination risk and data privacy concerns. Therefore, structuring AI project selection processes through Multi-Criteria Decision-Making (MCDM) methods grounded in analytical and mathematical principles has become a strategic necessity. The main objective of this study is to provide an integrated decision framework for identifying an AI project topic aligned with organizational strategic goals and selecting the most suitable LLM for that project. Within the scope of the study, a hierarchical structure encompassing strategic, technical, security, and cost dimensions was developed by integrating up-to-date criteria from the literature with expert opinions. To solve the decision problem, the Analytic Hierarchy Process (AHP), which enables the systematic and consistent quantification of expert judgments in complex decision-making settings, was adopted.

The research addresses three main questions: (1) Why is it important to carefully select both the topic of an AI project and the LLM to be used? (2) What key criteria should guide the selection of an AI project and an LLM? (3) Why is the AHP an appropriate methodology for this selection problem?

Following the introduction, the paper first presents a comprehensive literature review on AI project selection and LLM evaluation criteria, followed by a detailed explanation of the AHP methodology used to model the decision problem. The results section evaluates the alternatives according to the determined criteria and analytically presents the most suitable solution for organizational needs. The final section summarizes the main conclusions, discusses the limitations of the study, and provides recommendations for future research.

2. Literature Review

The abundance of LLM literature — over 38,000 Web of Science results for "LLM" OR "large language models" (retrieved November 2025) — itself presents a complex "choice paradox" for decision-makers. This section examines the strategic role of MCDM methods in managing this complexity and identifies the appropriate evaluation criteria from the most current studies.

2.1. AI Project Selection: Strategic Management Perspective

The rapid increase in the use of AI has led to an increase in AI-related studies in organizations, making project selection processes a critical decision-making area in terms of strategic management. AI projects involve high costs, high risks, and wide-ranging organizational impacts. Therefore, project selection is critical for the long-term strategic orientations of organizations.

Souza et al. (2021), by examining 263 criteria published over 50 years, grouped the main sets of criteria used in R&D and project selection under the headings of "strategic alignment, risk, cost, expected benefit, technical feasibility, and data requirement." While this study proves the multidimensional nature of project selection even in the pre-AI era, the content of these criteria has been expanded to include "data quality" and "algorithmic risks" in current AI projects.

Similarly, Kandakoglu et al. (2023) show that there is a multi-criteria environment in project portfolio selection (PPS) and that MCDM methods have become standard in PPS decisions. The research highlights the superiority of analytical methods over heuristic methods in improving decision quality, especially in project types involving high uncertainty and technological volatility, such as artificial intelligence projects. Wankhade et al. (2025) take this perspective a step further, stating that not only technical and financial aspects but also "ethical" and "social impact" dimensions (transparency, accountability) should be used as strategic filters in project selection.

2.2. MCDM Applications for AI Project Selection and Prioritization

A literature review reveals that the use of multi-criteria methods in project selection is becoming increasingly common. Methods such as FAHP (Fuzzy AHP), ANP, DEMATEL, and TOPSIS are particularly preferred for complex, risky, and multi-stakeholder projects.

Shaygan and Testik (2019) demonstrated that project selection under uncertainty can be done much more accurately with AHP-based methods. In the study, the advantages of the Fuzzy AHP model, which combines heuristic information sources and mathematical weighting for project selection, in quantifying uncertain statements, such as "high risk," were highlighted.

In 66 PPS articles examined by Souza et al. (2021), the most frequently used methods were identified as AHP, ANP, TOPSIS, VIKOR, and DEMATEL. This study argues that with the increasing use of MCDM in project selection models, multi-criteria methods are the correct choice in areas with high uncertainty, such as AI projects. In addition, Avramova et al. (2025), in their study on decision-making processes in industrial environments, stated that methods such as TOPSIS and VIKOR stand out in complex technology selection due to their computational efficiency. Danesh et al. (2017) further compared a range of MCDM methods for project portfolio management, confirming that no single technique dominates and that method choice must reflect the decision context.

2.3. Evaluation Criteria for Large Language Models (LLMs)

With the rapid spread of LLMs in corporate environments in recent years, the need to select the right application areas and the most suitable model has become a critical decision-making issue. However, since LLMs are emerging technologies, studies on LLM selection criteria are still relatively new in the literature and have intensified especially during 2024 and 2025.

Aksoy et al. (2025) evaluated LLMs according to five main trustworthiness dimensions using Hesitant Fuzzy AHP: fairness, robustness, integrity, explainability, and safety. Radulescu (2025) examined how criteria influence each other in LLM selection using a DEMATEL-based model and argued that cost, security, performance, usability, and community support are the main determinants. In particular, community support was found to directly affect the long-term performance of the model through a causal relationship. Akin et al. (2025) compared nine different LLM-supported assistants, such as ChatGPT, Gemini, and Claude, using the SAW method based on criterion weights determined with the G1 method. They argued that performance, user experience, ethical issues, and technical limitations are the most important criteria in the context of academic use. Alabool (2025) found that the criteria of reliability and robustness should be assigned much higher weights than cost-related criteria in the selection of LLMs in the healthcare sector.

2.4. Applications of MCDM in the Selection of LLM and Generative AI Models

Seiti et al. (2024) evaluated generative AI models for supply chain risk management by presenting an improved version of the EDAS method. This study is one of the pioneering works that proves that multi-criteria methods for large language models/generative AI model selection can be used from a human-centric perspective. Chen et al. (2025) combined the MARCOS method with Entropy and CRITIC weighting techniques in the selection of AI co-writing tools to be used in newsrooms. The study showed that "factual accuracy" and "workflow integration" are more critical criteria than creativity for the media sector. Aksoy et al. (2025) compared models in a multi-criteria framework using Hesitant Fuzzy AHP for reliability evaluation of LLMs. In the studies by Akin et al. (2025) and Radulescu (2025), it was shown that advanced methods such as TOPSIS, SAW, G1, and DEMATEL can be used in a hybrid manner in LLM selection. In a related vein, Bukar et al. (2024) proposed a decision-making framework for the adoption of generative AI in education, illustrating how structured MCDM reasoning extends to the deployment of tools such as ChatGPT. These studies indicate that model selection has now moved beyond being merely a technical preference and has become a managerial decision supported by analytical methods.

2.5. Summary and Evaluation

The literature review clearly reveals that the selection of artificial intelligence projects and the evaluation of Large Language Models (LLMs) inherently constitute a multi-criteria decision-making problem. Both in the project selection literature and in LLM evaluation studies, decision-makers are required to consider criteria from different dimensions that often conflict with one another. In this study, to address this multi-criteria decision-making need identified in the existing literature, the selection of the project topic and the LLM to be used is performed using the

AHP method. The hierarchical structure of AHP and its ability to calculate relative importance weights provide an analytical basis suitable for the complex and uncertain nature of artificial intelligence projects. Tables 1 and 2 present a comparative analysis of the methods and criteria sets used in the reviewed literature.

Table 1. Studies in the Literature and the MCDM Methods Used

Author / Year	Study Topic	Method(s) Used
Aksoy et al. (2025)	LLM Trustworthiness Evaluation	Hesitant Fuzzy AHP
Akin et al. (2025)	Comparison of Academic LLM Assistants	G1 (Weighting) + SAW (Ranking)
Alabool (2025)	LLM Selection in Healthcare	Fuzzy AHP
Chen et al. (2025)	GenAI Tool Selection for Newsrooms	Entropy, CRITIC (Weighting) + MARCOS (Ranking)
Radulescu (2025)	Criteria Relationship Analysis in LLM Selection	Improved Group DEMATEL (IDEMATEL)
Seiti et al. (2024)	GenAI Model Selection for Supply Chain	Improved EDAS
Wankhade et al. (2025)	Ethical AI Project Evaluation	Artificial Neural Networks (ANN) + ANFIS
Souza et al. (2021)	R&D Project Selection (Literature Review)	AHP, ANP, TOPSIS, VIKOR, DEMATEL
Shaygan and Testik (2019)	Project Prioritization	Fuzzy AHP
Avramova et al. (2025)	Industrial Decision-Making Methods	AHP, VIKOR, FUCOM, TOPSIS

Table 2. Key Evaluation Criteria Used in the Literature

Author / Year	Focus Area	Criteria Used
Souza et al. (2021)	Project Selection	Strategic Alignment, Risk, Cost, Expected Benefit, Technical Feasibility, Data Requirement.
Aksoy et al. (2025)	LLM Trustworthiness	Fairness, Robustness, Integrity, Explainability, Safety.
Akin et al. (2025)	Academic Assistant	Performance, User Experience (UX), Ethical Issues, Technical Constraints.
Alabool (2025)	Healthcare LLM	Reliability, Robustness, Bias, Cost, Usability.
Radulescu (2025)	General LLM	Cost, Security, Performance, Usability, Community Support.
Chen et al. (2025)	Media / Writing	Factual Accuracy, Workflow Integration, Speed, Language Coverage, Ethical Compliance.
Wankhade et al. (2025)	Project Ethics	Transparency, Accountability, Privacy, Social Impact.

According to evaluation of studies in the literature, summary of criteria groups are shown below:

- Managerial/Strategic: Strategic Alignment, Expected Benefit, Social Impact.
- Risk/Reliability: Risk, Resilience, Security, Robustness, Bias, Legal Compliance.
- Cost/Financial: Cost, ROI (Return on Investment), Financial Accuracy.
- Technical/Performance: Performance, Technical Feasibility, Scalability, Speed, Accuracy/Integrity.
- Ethical/Social: Fairness, Transparency/Explainability, Accountability, Ethical Issues, Privacy, Ethical Compliance.
- Implementation/Use: Usability, User Experience, Workflow Integration, Community Support, Learning Effect.

3. Methodology

This section details the research model and analysis methodology that form the basis of the study. Determining the AI project topic and selecting a suitable LLM constitute a multi-criteria decision problem with technical, financial, and ethical dimensions. To manage this complexity, the AHP method, one of the MCDM techniques, was chosen. The phases of the study are presented in Figure 1. The visual illustrates a systematic approach to decision-making based on AHP. The process begins with a detailed problem statement supported by a comprehensive literature review. The AHP methodology, involving matrix-based pairwise comparisons, is then implemented after a detailed evaluation of the selected criteria. To support an overall, data-driven selection, the process finishes with a well-informed analysis of the prioritized outcomes.

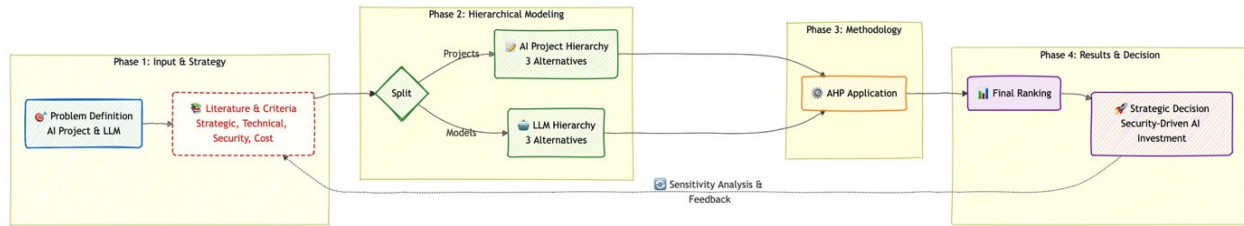


Figure 1. Phases of the Study

3.1. Method Selection: Analytical Hierarchy Process (AHP)

Developed by Saaty (1980), AHP structures complex decisions hierarchically, enabling the systematic prioritisation of both quantitative and qualitative expert judgements. Three literature-grounded reasons motivate its use here: first, AHP handles the high uncertainty of AI projects by converting verbal judgements such as “high risk” into numerical weights (Shaygan and Testik, 2019); second, it is among the most frequently used and empirically validated methods in R&D and project selection for more than five decades (Souza et al., 2021); and third, it ensures consistent relative weighting when criteria such as cost, reliability, and performance conflict, a critical property in LLM selection (Alabool, 2025).

3.2. Determining the Research Model and Criteria

The criteria to be used in the AHP analysis were derived both from the main dimensions identified through the comprehensive literature review and from the specific characteristics of the project, while also incorporating expert opinions in the field. The literature reveals that AI project selection and LLM evaluation should go beyond technical performance and encompass ethical and security-related dimensions. Accordingly, the criteria and supporting points used in the model developed for AI project selection are presented in Table 3, and the criteria and supporting points used in the model developed for LLM model selection are presented in Table 4. Two placements warrant a brief explanation. “Market Potential” is grouped under Cost and Efficiency because, in this study, it is treated as the revenue-side counterpart of cost — that is, the commercial return expected against the resources invested — rather than as a standalone strategic criterion. “Community Support” is grouped under Integration and Compatibility because an active developer and user community directly determines the availability of connectors, documentation, and maintenance that govern how readily a model can be integrated into existing systems.

Table 3. Proposed AHP Criteria and Sources (Project selection)

Main Criteria	Sub-Criteria	Description	Source
1. Strategic and Managerial Alignment	Strategic Alignment	Alignment of the project with the organization's long-term goals.	Souza et al. (2021)
	Corporate Integration	Technical compatibility of the selected project with existing workflows.	Expert Opinion
2. Technical Performance	Technical Feasibility and Constraints	What is the engineering effort, integration difficulty, and technological uncertainty level required by the project?	Souza et al. (2021)
	Human Resource Competency	Is the necessary expertise for the project available in-house?	Souza et al. (2021)
3. Security and Ethics	Data Security	Protection of sensitive data and leak risks.	Radulescu (2025)
	Legal Compliance	Ability to comply with legal regulations (e.g., GDPR/KVKK).	Souza et al. (2021)
4. Cost and Efficiency	Financial Return	What is the project's estimated Net Present Value (NPV) and Return on Investment (ROI)?	Radulescu (2025)
	Cost	Estimated cost calculation for the project.	Radulescu (2025)
	Market Potential	What is the market potential? What is the likelihood of market expansion and implementation?	Souza et al. (2021)

Table 4. Proposed AHP Criteria and Sources (LLM selection)

Main Criteria	Sub-Criteria	Description	Source
1. Integration and Compatibility	Data Privacy	Does the data need to be sent to the model provider (API) or can it be run in-house (On-premise)?	Radulescu (2025)
	Community Support	Developer community and ecosystem support for the model.	Radulescu (2025)
2. Technical Performance	Accuracy and Integrity	Hallucination rate and factual accuracy of responses.	Aksoy et al. (2025)
	Robustness	Resistance of the model against erroneous or manipulative inputs.	Aksoy et al. (2025)
3. Security and Ethics	Data Security	Protection of sensitive data and leak risks.	Radulescu (2025)
	Legal Compliance	Ability to comply with legal regulations.	M. Akin et al. (2025)
4. Cost and Efficiency	Cost	Training, API, and maintenance costs.	Radulescu (2025)
	<i>Fine-tuning Cost</i>	Cost of aligning the model with corporate data.	Expert Opinion

4. Application of the Proposed Methodology

This study was conducted as a single-organisation case study. The case organisation operates in the retail sector and is at an early-to-mid stage of AI adoption, characterised by experimentation and pilot implementation.. The three candidate projects (A1: Media and Content Management, A2: Campaign Analysis, A3: Supplier Assistant Bot) were drawn from the organisation's active AI agenda as the initiatives under live consideration, and the three candidate models (Llama, Gemini, GPT) were selected as the leading general-purpose LLMs evaluated for internal deployment, in line with Akin et al. (2025) and Aksoy et al. (2025).

4.1. Steps of the AHP Methodology

The AHP methodology used in this study will be applied in the following steps to provide a mathematical basis for decision-makers' subjective judgments and to enable them to make the most accurate decisions in uncertain situations:

Step 1: Defining the problem

The topic to be examined in the study is determined. What is the problem and how can it be solved? What criteria and alternatives should be considered?

Step 2: Modeling the Hierarchical Structure

The decision problem is broken down from top to bottom and transformed into a hierarchical tree structure. This structure consists of three basic levels (Figure 2 and Figure 3):

- Goal: To select the most suitable AI project topic or LLM model.
- Criteria: Main and sub-factors such as strategic alignment, integration and compatibility, technical performance, safety and ethics, cost and efficiency.
- Alternatives: Candidate projects to be evaluated were determined in line with the company's agenda and expert opinions. The selection of LLM models to be evaluated was determined in accordance with the sources M. Akin et al. (2025), Aksoy et al. (2025).
-

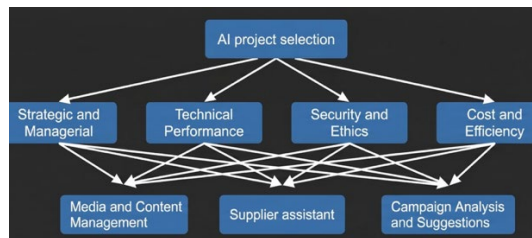


Figure 2. AI Project Selection AHP Modeling

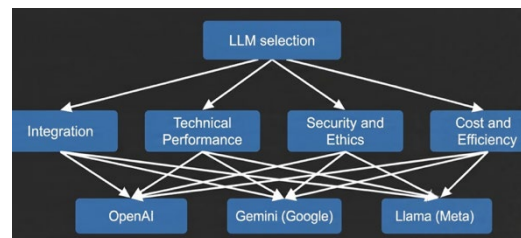


Figure 3. LLM Model Selection AHP Modeling

Step 3: Creating Pairwise Comparison Matrices

To evaluate the criteria and alternatives determined within the scope of the study, five experts with at least five years of experience in AI projects and LLM integration were consulted. The panel was composed of IT managers and business-development experts from the retail sector; none of the experts held a commercial affiliation with the LLM vendors under evaluation. The participants were recruited purposively from the case organisation on the basis of their direct involvement in AI adoption decisions. Because the entire weighting process is based on these five expert judgements, this composition defines the scope within which the findings should be interpreted. The experts were asked to conduct pairwise comparisons of the main criteria, sub-criteria, and alternatives using Saaty's 1–9 importance scale. Since individual judgements may differ, the geometric mean of the pairwise-comparison scores provided by the five experts was calculated to aggregate individual judgements into a single collective value. Based on these aggregated judgements, combined pairwise comparison matrices were constructed (Table 5).

Table 5. Comparison Scale (Saaty, 1980).

Degrees	Definition
1	equally as important as
3	moderately more important
5	strongly more important
7	very strongly more important
9	extremely more important

Step 4: Priority Vector Calculation

The eigenvector of the created comparison matrix is calculated to determine the weights (w) of the criteria.

$$A \cdot w = \lambda_{\max} \cdot w$$

Step 5: Priority Values Calculation

Priority values are calculated by the double comparison method. This step is the most important step used in the decision-making process. Comparisons are made in pairs to determine the priority order of the criteria or alternatives.

Step 6: Consistency Verification

The Consistency Ratio (CR) is calculated to test the logical consistency of expert judgments.

$$CI = \lambda_{\max} - n / (n - 1)$$

The generally accepted threshold in the literature is $CR < 0.10$. If this value is met, the decision is considered consistent. This validation process is essential for ensuring decision reliability in AI project selection involving high risk and uncertainty.

Step 7: Ranking of Decision Options

The priority vectors obtained for the criteria are combined to create a matrix for all criteria. A resultant vector is then formed from this matrix. The decision option with the highest weight in this vector is determined as the preferred decision option for solving the problem.

5. Results and Discussion

This section presents the final priority weights of the criteria and the resulting ranking of the alternatives. Subsequently, a comprehensive analysis is provided to interpret these outcomes and the optimal selections.

5.1 Numerical Results

Project Selection Criteria Weights (Global Weights)

The Global Weights were calculated by multiplying the main criterion weights by the local weights of the sub-criteria as a result of solving the matrices above. The results are summarized in Table 6 and Table 7.

Table 6. AI Project Selection Final Weight Table

Main Criteria	Weight (w)	Sub-Criteria	Local Weight	Global Weight	Ranking
Security and Ethics	0.47	Data Security	0.75	0.353	1
		Legal Compliance	0.25	0.117	3
Technical Performance	0.29	Technical Feasibility	0.67	0.194	2
		Human Resource	0.33	0.096	5
Strategic Alignment	0.17	Strategic Alignment	0.60	0.102	4
		Integration	0.40	0.068	6
Cost	0.07	Financial Return	0.54	0.038	7
		Cost	0.30	0.021	8
		Market Potential	0.16	0.011	9

Table 7. Consistency Ratios of the AI Project Selection Pairwise Comparison Matrices

Matrix	λ_{max}	CI	CR	Result
Main criteria	4.047	0.016	0.017	CR < 0.10
Sub-criteria: Strategic & Managerial Alignment	2.000	0.000	0.000	CR < 0.10
Sub-criteria: Technical Performance	2.000	0.000	0.000	CR < 0.10
Sub-criteria: Security & Ethics	2.000	0.000	0.000	CR < 0.10
Sub-criteria: Cost & Efficiency	3.006	0.003	0.005	CR < 0.10

LLM Selection Criteria Weights (Global Weights)

The final weighting table showing which factor is dominant in LLM model selection is given in Table 8 and Table 9.

Table 8. LLM Selection Final Weight Table

Main Criteria	Weight (w)	Sub-Criteria	Local Weight	Global Weight	Ranking
Security and Ethics	0.50	Data Security	0.67	0.335	1
		Legal Compliance	0.33	0.165	3
Technical Performance	0.31	Accuracy	0.60	0.186	2
		Robustness	0.40	0.124	4
Integration	0.12	Data Privacy	0.80	0.096	5
		Community Support	0.20	0.024	8
Cost	0.07	Usage Cost	0.60	0.042	6
		Fine-Tuning	0.40	0.028	7

Table 9. Consistency Ratios of the LLM Selection Pairwise Comparison Matrices

Matrix	λ_{max}	CI	CR	Result
Main criteria	4.038	0.013	0.014	CR < 0.10
Sub-criteria: Integration & Compatibility	2.000	0.000	0.000	CR < 0.10
Sub-criteria: Technical Performance	2.000	0.000	0.000	CR < 0.10
Sub-criteria: Security & Ethics	2.000	0.000	0.000	CR < 0.10
Sub-criteria: Cost & Efficiency	2.000	0.000	0.000	CR < 0.10

Ranking of AI Project Alternatives

The expert group scored the three identified potential project topics (A1: Media and Content Management, A2: Campaign Analysis, A3: Supplier Assistant Bot) using pairwise comparisons based on specific criteria. The resulting synthesis is presented in Table 10.

Table 10. AI Project Selection Final Synthesis and Ranking Table

Criteria	Global Weight	A1: Media and Content Management	A2: Campaign Analysis and Suggestions	A3: Supplier Assistant
Security and Ethics	0.470	0.28	0.15	0.57
Technical Performance	0.290	0.45	0.35	0.20
Strategic Alignment	0.170	0.55	0.30	0.15
Cost and Efficiency	0.070	0.40	0.40	0.20
TOTAL SCORE	1.000	0.384	0.246	0.370
RANKING		1	3	2

Ranking of LLM Alternatives

Three different Large Language Models (L1: Llama-Meta, L2: Gemini-Google, L3: OpenAI-GPT) were evaluated for use in the selected project. This ranking, where "Data Security" and "Accuracy" criteria were particularly decisive, is summarized in Table 11.

Table 11. LLM Model Selection Final Synthesis and Ranking Table

Criteria	Global Weight	L1: Llama (Meta - Open Source)	L2: Gemini (Google)	L3: OpenAI (GPT)
Security and Ethics	0.500	0.60	0.20	0.20
Technical Performance	0.310	0.20	0.35	0.45
Integration	0.120	0.25	0.40	0.35
Cost	0.070	0.50	0.30	0.20
TOTAL SCORE	1.000	0.427	0.279	0.295
RANKING		1	3	2

The analysis of the results is summarized below.

- **Security-Oriented Approach:** The most striking finding of the study is that the “Security and Ethics” criterion received the highest weight in both project selection (47%) and LLM selection (50%). This empirically supports the argument advanced by Wankhade et al. (2025) and Alabool (2025) that ethics, transparency, and trustworthiness have become a more critical strategic filter than technical capacity or cost in AI projects.
- **Cost Takes Second Place:** The relatively low weight assigned to the “Cost and Efficiency” criterion, approximately 7% in both decision models, indicates that organizations may be willing to tolerate higher financial costs when data security and appropriate technical infrastructure are at stake.
- **Project Selection Result:** Among the project alternatives, the “Media and Content Management” project was ranked first due to its strong alignment with the organization’s strategic goals and its manageable risk level. In contrast, the “Campaign Analysis” project, which requires intensive processing of customer data, received the lowest score due to data privacy risks.
- **Model Selection Result:** Among the LLM alternatives, the Llama (Meta) model ranked first despite lagging behind GPT in technical performance. This result can be attributed to its open-source structure, which enables on-premises deployment and provides an advantage in terms of data sovereignty. The finding is consistent with the decisive role of data privacy in model selection, as emphasized in the studies of Chen et al. (2025) and Radulescu (2025).

5.2. Sensitivity Analysis

Because the “Security and Ethics” criterion dominates both decision models, a one-way sensitivity analysis was performed to test how robust the final rankings are to changes in this weight. The weight of the Security and Ethics criterion was perturbed by $\pm 10\%$ and $\pm 20\%$ around its baseline value (47% for project selection and 50% for LLM selection), and the weights of the remaining criteria were renormalised proportionally. The resulting top-ranked alternatives for each scenario are summarised in Table 12.

Table 12. One-Way Sensitivity Analysis of the Security and Ethics Criterion Weight

Scenario	Security & Ethics weight (Project / LLM)	Top-ranked project	Top-ranked LLM
Baseline	47% / 50%	A1 – Media & Content	L1 – Llama (Meta)
–20%	37.6% / 40%	A1 – Media & Content	L1 – Llama (Meta)
–10%	42.3% / 45%	A1 – Media & Content	L1 – Llama (Meta)
+10%	51.7% / 55%	A3 – Supplier Assistant	L1 – Llama (Meta)
+20%	56.4% / 60%	A3 – Supplier Assistant	L1 – Llama (Meta)

The sensitivity analysis reveals an asymmetric pattern. For the LLM selection, L1 (Llama) remains the top-ranked alternative across all scenarios, from a 20% decrease to a 20% increase in the Security and Ethics weight, with its score increasing from 0.392 to 0.462. This confirms that the LLM ranking is robust to reasonable variation in the dominant criterion. For the AI-project selection, however, the ranking is not fully robust: A1 (Media and Content Management) leads at the baseline weight of 47% and under downward perturbations (0.402 at 37.6% and 0.393 at 42.3%), but a rank reversal occurs when the weight is increased. At +10% (51.7%) and +20% (56.4%), A3 (Supplier Assistant) overtakes A1, with the two alternatives crossing over just above the baseline weight. The project recommendation should therefore be treated as conditional on the relative emphasis placed on Security and Ethics, whereas the LLM recommendation is stable.

6. Conclusion

The findings confirm that “Security and Ethics” is the dominant factor in both AI project selection (47%) and LLM selection (50%), supporting the view of Wankhade et al. (2025) and Alabool (2025) that trustworthiness and ethical compliance have become more decisive strategic filters than cost or technical performance. In practice, “Media and Content Management” ranked first due to its strategic alignment and manageable risk profile, while “Campaign Analysis” ranked last owing to data privacy risks. In LLM selection, Llama (Meta) prevailed despite lower raw performance, driven by its advantages in data sovereignty and on-premises deployment.

Overall, the study provides case-based evidence that AI investment success depends on balancing strategic alignment, data security, and technical requirements within a coupled decision framework, rather than simply selecting the most technically advanced solution.

6.1. Proposed Improvement

Based on the analytical results obtained, the following strategic steps are recommended for organizations undergoing artificial intelligence transformation. These recommendations should be understood as general best-practice guidance informed by the AHP analysis, rather than as direct outputs of it:

- **Data Sovereignty Strategy:** Instead of sending sensitive data to external closed-source models via APIs, organizations should consider using open-source models, such as Llama, that can be hosted on their own servers. This approach can help minimize GDPR-related and data security risks.
- **Hybrid Architecture:** Organizations should design a hybrid architecture in which high-performance models, such as GPT, are used for creative tasks with low data sensitivity, such as content creation, while locally hosted models are used for processes requiring internal data analysis.

- **Ethical Control Mechanism:** Since security emerged as the most important criterion in the study, organizations should establish a human-in-the-loop mechanism to monitor the accuracy of AI outputs, including hallucination control, as well as potential bias.

6.2. Limitations of the Study and Suggestions for Future Research

Several limitations should be acknowledged. First, the analysis is based on a relatively small expert panel of five participants drawn from a single case organisation; while this panel size is consistent with prior AHP applications and the experts satisfy the experience threshold defined in Section 4, it constrains the statistical generalisability of the findings to other sectors and organisational contexts. Second, the present study reports the consistency ratio (CR) of the aggregated judgements but does not include a formal sensitivity analysis of the criteria weights; the supplementary scenario analysis in Section 5.2 partially addresses this limitation by testing the robustness of the rankings to changes in the weight assigned to the dominant Security and Ethics criterion, yet a fuller variance-based sensitivity analysis remains outside the present scope. Third, this study is based on the judgements of a group of experts from a specific sector, and the AHP method used in the study requires experts to provide crisp judgements. However, human reasoning generally involves uncertainty, ambiguity, and hesitation. Therefore, future studies may extend the proposed framework by using methods such as Fuzzy AHP or Hesitant Fuzzy AHP to better reflect the linguistic uncertainty and hesitation of decision-makers. In addition, the Analytic Network Process (ANP) or DEMATEL methods may be incorporated into the model to capture the interdependencies among criteria, such as the relationship between cost and performance. The analysis may also be repeated with a more comprehensive set of alternatives by including additional models such as Claude and Mistral.

In conclusion, this study has analytically demonstrated that the success of AI projects is not merely about choosing the most technically advanced algorithm, but rather about establishing the right balance between corporate strategy, data security, and technical requirements.

Acknowledgments

The authors gratefully acknowledge the experts whose insightful contributions significantly enriched this research. Their guidance and expertise were essential to the successful completion of this study. This study is financially supported by the Galatasaray University Research Fund (FBA-2026-1346).

References

- Akin, M., Batur Sir, G. D., Karadağ, A. A. and Çerçioğlu, H., A multi-criteria comparison of large language model powered assistants in pre-research studies for academia, *IEEE Access*, vol. 13, pp. 127098–127115, 2025.
- Aksoy, M., Adem, A. and Dağdeviren, M., Trustworthiness evaluation of large language models using multi-criteria decision making, *IEEE Access*, vol. 13, pp. 168183–168200, 2025.
- Alabool, H. M., Large Language Model Evaluation Criteria Framework in Healthcare: Fuzzy MCDM Approach, *SN Computer Science*, vol. 6, no. 1, pp. 57, 2025.
- Avramova, T., Peneva, T. and Ivanov, A., Overview of Existing Multi-Criteria Decision-Making (MCDM) Methods Used in Industrial Environments, *Technologies*, vol. 13, pp. 444, 2025.
- Bukar, U. A., Sayeed, M. S., Razak, S. F. A., Yogarayan, S. and Sneesl, R., Decision-Making Framework for the Utilization of Generative Artificial Intelligence in Education: A Case Study of ChatGPT, *IEEE Access*, vol. 12, pp. 95368–95389, 2024.
- Chen, F., Bulgarova, B. A. and Kumar, R., Prioritizing Generative Artificial Intelligence Co-Writing Tools in Newsrooms: A Hybrid MCDM Framework for Transparency, Stability, and Editorial Integrity, *Mathematics*, vol. 13, no. 23, pp. 3791, 2025.
- Danesh, D., Ryan, M. J. and Abbasi, A., A systematic comparison of multi-criteria decision making methods for the improvement of project portfolio management, *International Journal of Management and Decision Making*, vol. 16, no. 3, pp. 281–315, 2017.
- Kandakoglu, M., Walther, G. and Ben Amor, S., The use of multi-criteria decision-making methods in project portfolio selection: A literature review and future research directions, *Annals of Operations Research*, vol. 332, pp. 807–830, 2023.
- Li, Y., Progress, applications, and challenges of large language models, *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE)*, SciTePress, pp. 645–648, 2025.

- Radulescu, C. Z. and Radulescu, M., Criteria analysis for the selection of a generative artificial intelligence tool for academic research based on an improved group DEMATEL method, *Applied Sciences*, vol. 15, no. 10, pp. 5416, 2025.
- Saaty, T. L., *The Analytic Hierarchy Process*. McGraw-Hill, New York, 1980.
- Seiti, H., Javadi, R., Ghanbari, H., Kazemiyan, S. K. and Hameed, I. A., Unleashing the potential of human-centric generative AI in supply chain risk management: A causal MCDM approach, *SSRN*, 2024.
- Shaygan, A. and Testik, Ö. M., A fuzzy AHP-based methodology for project prioritization and selection, *Soft Computing*, vol. 23, pp. 1309–1319, 2019.
- Souza, D. G. B., dos Santos, E. A., Soma, N. Y. and da Silva, C. E. S., MCDM-Based R&D Project Selection: A Systematic Literature Review, *Sustainability*, vol. 13, pp. 11626, 2021.
- Şimşek, B., Analysis of generative AI technologies' adoption using interpretive structural modeling: Empirical findings from small and medium-sized IT enterprises in Türkiye, M.S. Thesis, Graduate School of Informatics, Middle East Technical University, 2025.
- Wankhade, S., Sahni, S., León-Castro, E. and Olazabal-Lugo, M., Navigating AI ethics: ANN and ANFIS for transparent and accountable project evaluation amidst contesting AI practices and technologies, *Frontiers in Artificial Intelligence*, vol. 8, pp. 1535845, 2025.

Biographies

Gül Mine Ak is an MSc student in Industrial Engineering at Galatasaray University, Istanbul. She received her B.S. degree in Industrial Engineering at Istanbul Technical University. Her research interest includes optimization, supply chain management, artificial intelligence, MCDM techniques, strategic management and project management. In addition, she works at the Research and Development & Business Development Department of a retail company.

Gülçin Büyüközkan is an academic in Industrial Engineering at Galatasaray University, where she has been serving as a full professor since 2009. She received her undergraduate degree from Istanbul Technical University and completed her graduate and doctoral studies in Industrial Engineering at Boğaziçi University and the Académie de Grenoble, France. Her research expertise lies in multi-criteria decision-making (MCDM), fuzzy logic, artificial intelligence, and decision support systems. She has made significant contributions to the development of advanced fuzzy MCDM methodologies and their applications across diverse domains, including supply chain management, digital transformation, sustainability, smart cities, energy systems, and healthcare sector.