Journal of Business Decision and Risk Sciences (JBDRS)

https://doi.org/

Journal Homepage: www.JBDRSJournal.com

Journal of Business Decision and Risk Sciences

ORIGINAL RESEARCH PAPER

A Comparative Study between Artificial Intelligence and Human Ability in Implementing Project Management Knowledge Areas

Ali Jooyafar; Mohammad Ali Hatefi *; Majid Shakhsi-Niaei; Mohammad Reza Bahram-Choobin

Department of Petroleum Economics and Energy Diplomacy, Tehran Faculty of Petroleum, Petroleum University of Technology (PUT), Abadan, Iran.

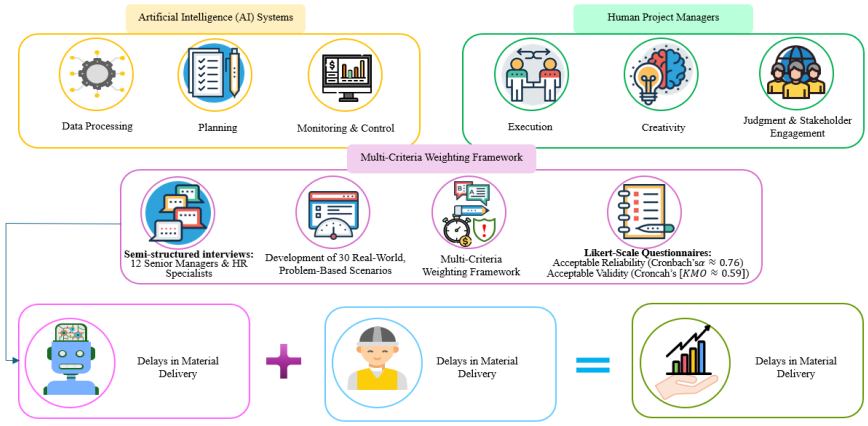
P A P E R I N F O

Paper history:
Received: 07/02/2026
Revised: 08/15/2026
Accepted in revised form: 08/25/2026
Published: 09/30/2026

***Keywords*:**
AI;
PM Knowledge Areas;
Iran Oil and Gas Industry;
Downstream Sector.

A B S T R A C T

This study presents a comparative analysis of human project managers and Artificial Intelligence (AI) systems in applying Project Management knowledge areas across different project phases in Iran's downstream oil and gas sector. Data were collected through semi-structured interviews with 12 senior managers and human resource specialists, resulting in the development of 30 real-world, problem-based scenarios. The responses generated by human managers and AI systems were subsequently assessed by professional evaluators using a multi-criteria weighting framework that included cost, time, risk, quality, and communication. Evaluation was conducted through Likert-scale questionnaires. The research instruments demonstrated acceptable reliability, with a Cronbach's alpha coefficient of 0.76, and acceptable validity, with a Kaiser–Meyer–Olkin (KMO) value of approximately 0.59. The findings indicate that AI performs at a level comparable to human managers in structured and data-driven project phases, particularly Planning, Monitoring, and Control. Nevertheless, human managers outperform AI in Execution and Closure, where creativity, contextual judgment, stakeholder engagement, and adaptive decision-making are particularly important. Overall, the findings support the development of a hybrid human–AI Project Management model that combines AI's algorithmic precision with human adaptability to enhance project efficiency, resilience, and overall success.



1. Introduction

This chapter establishes the foundation for a comparative investigation of Artificial Intelligence (AI) and human capabilities in Project Management (PM) within Iran's downstream oil and gas industry. It presents the sectoral context, the limitations of traditional PM approaches, and the transformative potential of AI. The downstream industry, which encompasses crude-oil refining and end-product distribution, is globally recognized for its complexity and its stringent requirements for efficiency, schedule compliance, safety, and environmental performance. Although human Project managers traditionally rely on expertise, intuition, and strategic judgment, growing project complexity and data volumes require greater precision, speed, and consistency [1].

Recent AI developments, particularly Gemini 2.5 pro 06-05, O3 Pro (Reasoning), O4-mini-high, GPT-5 Pro, Grok 3 Thinking, Claude 3.7 Sonnet Thinking, Claude Opus 4 Thinking, DeepSeek R1, and QwQ 32B, have created new opportunities for automating and improving PM activities. These systems

* Corresponding Author: Mohammad Ali Hatefi
E-mail address: hatefi@put.ac.ir
ISSN:
URL: https://www.jbdrsjournal.com/article_735690.html

Please cite this article as: J. Ali, M. A. Hatefi, & M. Shakhsi-Niaei, & M. R. Bahram-Choobin, *Journal of Business Decision and Risk Sciences (JBDRS)*, 2026, 2(3), 103-111. https://doi.org/

/© 2026 The Author(s). Published by Power of New Vision Movement Institute.

can process extensive datasets, optimize resource allocation, and generate predictive insights; nevertheless, their ability to resolve conflicts, manage stakeholder expectations, address ethical considerations, and respond to unforeseen circumstances remains uncertain [2,3]. Accordingly, this study examines whether AI can complement, enhance, or replace human expertise in managing critical downstream oil and gas projects.

The problem arises from the increasing difficulty of using conventional PM methods to address modern projects' data demands, time constraints, operational risks, regulatory requirements, and culturally sensitive decisions [4]. While Chatbots can rapidly analyze data, generate optimized solutions, and support critical management during project phases, their capacity for strategic planning, stakeholder engagement, and nuanced judgment requires further examination [5]. Using scenario-based and situational PMP questions aligned with PMBOK guidelines, this research compares AI and experienced human Project managers across the five phases of the project life cycle to determine where each performs most effectively.

The study is significant because it addresses persistent challenges, including project delays, cost overruns, and operational inefficiencies, while supporting more resilient PM practices in Iran's downstream oil and gas sector [6,7]. It evaluates Chatbot capabilities in data-driven tasks, strategic decision-making, risk management, scheduling, cost estimation, and stakeholder management. The research is guided by three questions: 1. Which PM areas defined by PMBOK are more suitable for AI, and where do human managers retain an advantage? 2. How does AI perform in strategic tasks compared with routine, data-intensive tasks? 3. What are the practical limitations and opportunities for integrating AI into traditional PM practices?

The oil and gas industry is divided into three sectors, namely upstream, midstream, and downstream. The upstream sector focuses on exploration and production, the midstream sector on transportation and storage, and the downstream sector on refining, distribution, and marketing. Downstream products include gasoline, diesel, heating oil, lubricants, and chemicals. High investment requirements, long project cycles, stringent environmental regulations, and geopolitical influences are among the principal characteristics of this industry [8–9]. This framework is illustrated in Figure 1 Fishbone Diagram of Oil and Gas Industry Challenges.

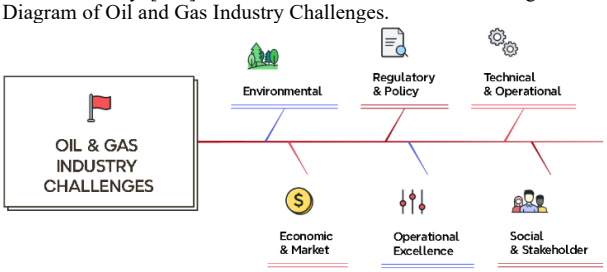


Figure 1. Fishbone diagram of oil and gas industry challenges

Iran's oil and gas industry faces international sanctions, limited access to technology, reduced foreign investment, and difficulties in connecting to the global supply chain [10]. Oil price fluctuations and economic instability also increase financial uncertainty in project planning and implementation. Aging infrastructure, restricted access to advanced drilling and refining technologies, and shortages of skilled personnel have reduced productivity. Bureaucratic inefficiency, regulatory restrictions, lengthy licensing procedures, additional costs, and environmental concerns are among the other challenges [11]. Downstream projects are highly complex because of their large scale, high capital intensity, interdependence among engineering, logistical, and regulatory factors, and the need for coordination among government bodies, contractors, suppliers, and investors [9,11-13].

Project Management is the application of knowledge, skills, tools, and techniques to achieve project objectives within the constraints of time, budget, quality, and scope. In the downstream oil and gas industry, it is essential for coordinating teams, optimizing resources, controlling costs, managing risks, complying with regulations, and meeting stakeholder expectations [14,15]. PMBOK, the recognized standard of the PMI, encompasses 10 knowledge areas, including Integration, Scope, Schedule, Cost, Quality, Resource, Communication, Risk, Procurement, and Stakeholder Management [16]. The PMP certification also confirms an individual's competence in applying PMBOK methods and best practices [17]. Situational PMP questions assess decision making, problem solving, evaluation, management, and ethical considerations under real world conditions [18,19]. The five PMBOK process groups are Initiating, Planning, Executing, Monitoring and Controlling, and Closing [9,20–25], as presented in Figure 2 Process Groups Flow.



Figure 2. Process groups flow

The relationship between knowledge areas and process groups is presented in Table 1 Knowledge Areas and Their Relations to Process Groups [26].

Table 1. Knowledge Areas and Their Relations to Process Groups

Knowledge Area	Relevant Process Groups
Integration	Initiating, Planning, Executing, Monitoring and Controlling, Closing
Scope, Schedule, Cost	Planning, Monitoring and Controlling
Quality, Resource, Communication, Risk, Procurement	Planning, Executing, Monitoring and Controlling
Stakeholder Management	Initiating, Planning, Executing, Monitoring and Controlling

Source: Ref [26]

Leadership is based on inspiration, trust building, effective communication, vision, empathy, and decision making under pressure [27,28], whereas management focuses on planning, resource coordination, progress control, time, budget, and risk [29–32]. The combination of leadership and management is essential for project success [33]. Decision making and planning also complement each other through selecting alternatives and determining objectives, tasks, resources, schedules, and risk mitigation strategies [32,34–37]. Problem solving, strategic thinking, and critical thinking are likewise important for analyzing obstacles, anticipating trends, and evaluating assumptions and solutions [38–41].

The evolution of AI can be traced from the McCulloch–Pitts artificial neuron model in 1943, the Turing Test in 1950, the emergence of ELIZA in 1967, the AI winter from the 1980s to the early 1990s, and the increase in AI use for image generation to 70% in 2023. The term AI was introduced by John McCarthy in 1955, and the field expanded through the development of expert systems, machine learning, Deep Blue's victory over Garry Kasparov in 1997, and the growth of Deep Learning in the 2010s [42–46]. AI is used for pattern recognition, language understanding, problem solving, and prediction, encompassing areas such as ML, Deep Learning, and NLP [42,43]. Figure 3 Stack Venn Chart of AI Concepts, Figure 4 Expert System Diagram Designed with GPT-4o-Image Model, Figure 5 The Turing Test Diagram, and Figure 6 Turing Test Limits illustrate the core concepts of this section. Searle's Chinese Room experiment in 1980 also highlights the distinction between syntactic processing and genuine semantic understanding [47-51].

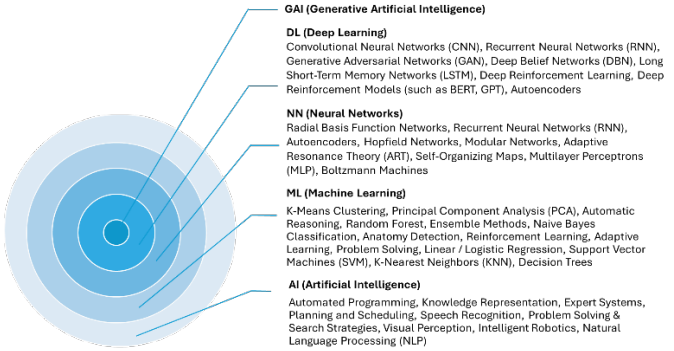


Figure 3. Stack Venn Chart of AI Concepts

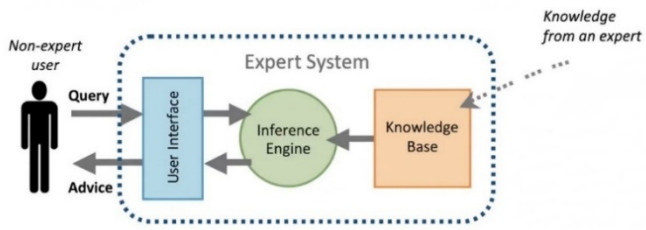


Figure 4. Expert system diagram designed with GPT-4o-image model

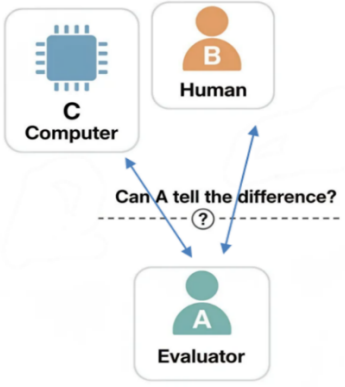


Figure 5. The turing test diagram

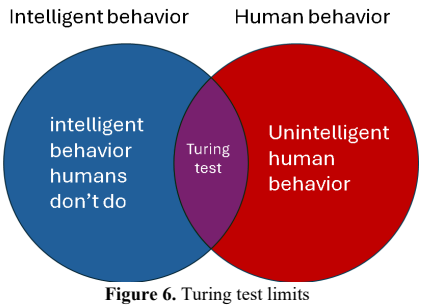


Figure 6. Turing test limits

NLP is used for understanding and generating human language in Chatbots, translation, sentiment analysis, voice assistants, and summarization. However, it faces limitations in interpreting sarcasm, slang, and ambiguous language [52]. NLU focuses on semantic language understanding and includes syntax analysis, semantics, entity recognition, intent recognition, and context understanding [53,54]. A neural network consists of Input, Hidden, and Output layers and adjusts its weights through training and backpropagation to reduce errors [55–58]. Figure 7 Three Types of Layers in Neural Networks, Generated by the GPT-4o Image Model presents this structure. GPT, a Transformer based language model, generates text by predicting the next word, and its GPT-2, GPT-3, and GPT-4 versions, despite containing billions of parameters, still rely on probabilistic patterns rather than genuine understanding [59–61]. Chatbots are divided into rule based and AI driven types [62–64]. In a test involving 250 prompts, the phrase "very rudely" achieved 84.8% accuracy, whereas "very polite" achieved 80.8% accuracy [65]. The AUTOMAT strategies include Role Definition, Contextual Framing, Instruction Structuring, Exemplar Inclusion, Reasoning Staging, Iterative Adjustment, and Output Specification. Figure 8 Conceptual Example for AUTOMAT-Prompt Principles [66,67].

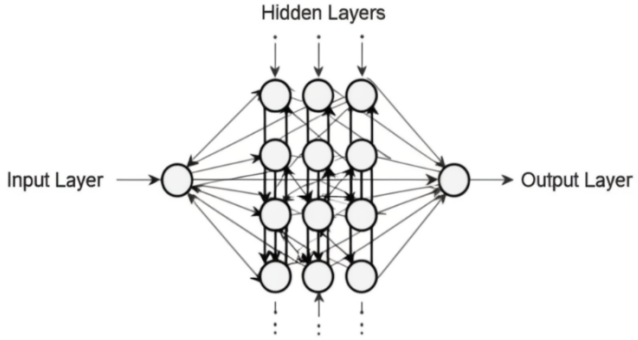


Figure 7. types of layers in neural networks, generated by the GPT-4o image model

So the arrows' directions and heads may not be accurate as graphically.



Figure 8. Conceptual Example for AUTOMAT-Prompt Principles

AI is capable in data analysis, automation, and prediction. However, compared with humans, it is weaker in emotional intelligence, creativity, leadership, conflict resolution, team motivation, and stakeholder interaction [68–72]. Poor or incomplete data quality, encountering new information, weaknesses in ethical judgment, and limitations in long term strategic thinking are among the principal challenges of AI. Cultural factors, data bias, and differences in communication norms may also generate inappropriate recommendations and increase the need for data diversity and localization [73-76]. Algorithmic bias, data privacy and security, transparency, explainability, and accountability for errors are among the most important ethical considerations [30,77,78], while job displacement requires employee retraining [79]. Conversely, human managers face fatigue, time and information constraints, and cognitive biases, whereas AI has greater processing speed and scale [80–86]. The concepts of EQ [87,88], confirmation, anchoring, and overconfidence biases [89], and bounded rationality [90,91] are important for explaining these differences.

Human knowledge extraction is conducted to convert the tacit knowledge of managers and teams into explicit knowledge [92,93], encompassing methods such as interviews and surveys, knowledge mapping, collaborative platforms, workshops, focus groups, and documentation [94,95]. Table 2, Knowledge Extraction Methodologies [92] compares four types of interviews, namely Structured, Semi-Structured, Unstructured, and Focus Groups, as well as the techniques of 5W1H, Laddering, Critical Incident, Think-Aloud, Scenario-Based Questioning, and Devil's Advocate. Questionnaires face response bias, sampling issues, limited depth of responses, misunderstanding, low response rates, privacy concerns, validity, reliability, and respondent fatigue [96,97]. Interviews also face bias, high time and cost requirements, limited sample sizes, access difficulties, and analytical complexity [98–100]. Studies indicate that AI is effective in Cost, Scheduling, and Risk Management, but has less impact on stakeholder interaction and team management [101–105]. Maheshwari's study in 2025 increased CAPEX forecasting accuracy from 65–75% to 85–92%, halved cost overruns, and reported an ROI of up to 6× [106]. Domestic studies also emphasize risk optimization, improved efficiency, and human–AI collaboration [107,108]. Ultimately, the proposed hybrid model suggests that AI should perform repetitive tasks while managers focus on strategic and complex decisions. AI training and an ethical framework are also necessary for this integration [104-110].

2. Materials and Methods

This study utilizes a scenario-based comparative framework to evaluate the project management efficacy of leading AI reasoning models—including Gemini 2.5 Pro 06-05, O3 Pro, O4-mini-high, GPT-5 Pro, Grok 3, Claude 3.7 Sonnet, Claude Opus 4, DeepSeek R1, and QwQ 32B—against experienced human managers within Iran's downstream oil and gas sector. By anchoring real-world challenges in the PMBOK Guide across the five project life-cycle phases, the research assesses whether advanced deep learning architectures can rival human expertise in critical areas like decision-making, scheduling, and risk management. Motivated by the sector's mounting operational and regulatory complexities, this study builds upon the iterative evolution of large language models, specifically leveraging the enhanced contextual reasoning of architectures like GPT-4 to provide evidence-based recommendations for integrating AI while preserving essential human leadership and stakeholder engagement.

To replicate the evaluative rigor of the PMP certification process and the practical realities of downstream project management, the scenarios cover all five PMBOK-defined process groups—initiation, planning, execution, monitoring and control, and closure—and are aligned with the ten PMBOK knowledge areas: Integration, Scope, Time, Cost, Quality, Human Resource, Communication, Risk, Procurement, and Stakeholder Management. The tasks are designed to elicit procedural knowledge, data analysis, scheduling, cost estimation, risk evaluation, contextual decision-making, critical management, and stakeholder engagement, thereby reflecting the operational, regulatory, and managerial demands of Iran's downstream oil and gas sector. The assessment follows an iterative questioning protocol, moving from broad questions to more specific prompts that require strategies, explanations, assumptions, and justifications, and responses are independently evaluated by HR experts using a standardized 1-to-5 Likert scale across Decision Quality, Risk Management, Stakeholder Management, Strategic Thinking, Efficiency, Accuracy, and Communication. This quasi-experimental, scenario-driven design presents identical realistic PM scenarios to human and AI participants, all derived from PMBOK guidelines and adapted to the downstream context, to compare the strengths of G-AI models, identify domains where human expertise remains indispensable, and determine when hybrid human–AI collaboration is most effective. The human arm includes nine PMP-certified project managers with≥7 years of downstream oil and gas experience, selected through purposive sampling based on the sector's roughly 35-year career span and a Pareto-based assumption that 7 years of experience captures about 20% of career time while covering nearly 80% of essential operational requirements; additionally, the expert panels comprised 12 experts in Stage 1 interviews and 11 experts in Stage 2 evaluations, all with ≥5 years of downstream project management experience and strong familiarity with PMBOK and industry assessment practices.

The AI arm consists of nine advanced reasoning-capable large language models, as presented in Table 1 These models were selected purposively because of their state-of-the-art natural language understanding, reasoning, contextual interpretation, and decision-support capabilities across the PMBOK process groups. The diversity of the selected systems provides representation across multiple developer ecosystems, model architectures, training approaches, and reasoning configurations. This design increases the representativeness of the AI arm and reduces dependence on the characteristics of a single model family.

Table 1. AI reasoning chatbots (Arms)

Model	Description
Gemini 2.5 Pro (06-05)	Google DeepMind's multimodal flagship model, based on the June 2025 build, with enhanced reasoning, coding, and visual capabilities and an orientation toward enterprise and research-grade conversational accuracy.
O3 Pro (Reasoning)	OpenAI's reasoning-focused model, designed for complex analytical tasks, chain-of-thought processing, and logic-intensive decision support.
O4 Mini High	A lightweight GPT 4 class variant offering high efficiency, low latency, and balanced accuracy for prompt engineering and rapid iteration tasks.
GPT 5 Pro	OpenAI's 2025 premium large language model, offering advanced multimodal reasoning, extensive context retention, and high precision across disciplines.
Grok 3 Thinking	X (AI)'s third-generation reasoning model, optimized for real-time discourse, creative synthesis, and adaptive problem solving across dynamic topics.
Claude 3.7 Sonnet Thinking	Anthropic's balanced model, combining substantial reasoning depth with moderate speed and suitability for professional writing and critical analysis.
Claude Opus 4 Thinking	Anthropic's most powerful series variant, featuring extensive context windows and nuanced reasoning for long-form, knowledge-dense tasks.
DeepSeek R1	An open Chinese-developed LLM emphasizing resource efficiency, scalable fine-tuning, and robust multilingual analytical capabilities.
QwQ 32B	A 32 billion-parameter community model emphasizing open-source adaptability, high inference stability, and reasoning performance comparable to larger proprietary systems.

To ensure experimental equity, both study arms receive the same set of 30 refined project management scenarios generated in Stage 1 and filtered in Stage 2, distributed equally with six scenarios assigned to each of the five PMBOK process groups. Consequently, the human arm produces 270 raw responses (5 phases × 6 scenarios × 9 project managers) and the AI arm produces 270 raw outputs (5 phases × 6 scenarios × 9 models), guaranteeing that performance differences are attributed to respondent type rather than unequal task conditions. To manage evaluator workload and normalize intra-group variation, all responses of the same type for a given scenario are merged into a single "combined human answer" and a "combined AI answer," reducing the evaluation volume per HR expert from 2,700 individual judgments (5 parameters × 6 scenarios × 5 phases × 9 respondents × 2 arms) to 300 judgments (5 parameters × 6 scenarios × 5 phases × 1 merged answer × 2 arms). This merging process preserves parameter-level content and substantive information while eliminating repetition, establishing a balanced, controlled, PMBOK-compliant, and operationally feasible framework for comparing decision-making capabilities in Iran's downstream oil and gas sector [111, 112].

2.1. Data Collection Procedures

The research employs a four-stage, quasi-experimental, scenario-based design to collect comparable decision-making responses from human and AI participants. The procedures are structured to preserve methodological rigor, ensure contextual relevance, and reproduce the operational conditions of Iran's downstream oil and gas Project Management environment. The four stages include scenario generation and criteria definition, scenario filtering and parameter selection, response collection from the two study arms, and evaluation by independent HR experts.

Table 2. Data Collection Procedures

Stage	Focus	Key Activities	Output
1	**Scenario & Criteria Definition**	Expert interviews (HR & PM) to generate >30 scenarios testing managerial competence; identifying & weighting criteria (1–5).	>30 initial scenarios; weighted evaluation parameters.
2	**Filtering & Selection**	Independent expert panel scores scenarios on importance × repeatability; selects top 6 per PMBOK group; selects top 5 criteria.	30 refined scenarios; final 5-parameter rubric.
3	**Response Collection**	9 PMP-certified managers vs. 9 high-end AI models respond to 30 scenarios; merging responses by "arm" to normalize data.	30 "combined human" vs. 30 "combined AI" responses.
4	**Comparative Evaluation**	Independent HR evaluators rate paired responses using an 11-point Likert scale (1=Human > AI, 6=Equal, 11=AI > Human).	300 comparative assessments for statistical analysis.

2.2. Scenario Generation and Filtering

During the initial stage, seven HR professionals and Project Management experts from Iran's downstream oil and gas sector were asked to identify real-world challenges commonly presented to Project Management applicants. A total of 134 distinct scenario-based challenges were collected and distributed across the five phase groups corresponding to the PMBOK process areas, with each group containing between 25–27 initial cases. AI tools were then used to assign concise titles and verify the independence of the scenarios by identifying duplications and conceptual overlaps. Subsequently, 23 additional experts evaluated every scenario in terms of repeatability, time/cost intensity, and importance using a 1 to 5 scale.

The final score of each scenario was calculated as follows:

$$ScenarioScore_i = Mean\ (Repeatability_i) \times Mean(Time - Cost_i)$$

The six scenarios with the highest scores in each phase group were retained as representative cases. Where two scenarios were dependent or substantially overlapping, the dependent scenario was replaced by the next-ranked independent alternative. This filtering process produced 30 final independent challenge scenarios distributed equally across the five PMBOK process groups.

2.3. Performance Rating Scale

In Stage 4, independent HR evaluators assessed each paired merged answer for every retained parameter through an 11-point quasi-Likert scale. The scale was defined as 1 = Human Much Better, 6 = Equal Performance, and 11 = AI Much Better, while the intermediate integers 2–5 and 7–10 captured different degrees of relative performance. Eleven evaluators, comprising scenario designers and Project Management experts, assessed the paired human and AI responses against the five most important criteria identified during the earlier expert-scoring process. To preserve symmetry, the combined value of the human and AI scores for each row was fixed at 12, enabling balanced comparative analysis.

The final composite score for each scenario was calculated using the following hybrid multi-criteria formula:

$$ScenarioScore_i = Mean\ (EvaluatorScore_i) \times ScenarioScore_i \times Mean(CriteriaImportance_i)$$

This formula integrates the characteristics of the scenario, the judgment of the evaluators, and the relative importance of the evaluation criteria. Consequently, it provides a unified weighted decision metric for comparing the two participant arms across the five project phases.

2.4. Weighted Scoring Formula

For scenario s and parameter i , the raw evaluator score was multiplied by the parameter weight. The weighted score was expressed as follows:

$$WeightedScore_{(si)} = Score_{(si)} \times Weighted_i$$

The total weighted score of scenario s was then calculated as:

$$ScenarioScore_s = \sum_{i=1}^{5} WeightedScore_{(si)}$$

Because the scenarios were evaluated according to multiple criteria with different levels of importance, a weighted scoring methodology was used to aggregate expert judgments into a single quantitative measure. The weighted score for scenario i, assessed by judge j, was calculated as:

$$S_{ij} = \sum_{k=1}^{K} w_k r_{ijk} \quad (1)$$

Here, S_{ij} represents the weighted score for scenario i as evaluated by judge j, and w_k denotes the weight assigned to parameter k, with $\sum w_k = 1$ The weights were predefined according to pilot findings or theoretical importance so that the total weight of all parameters equaled unity, or 100%. The term r_{ijk} represents the raw rating assigned by judge j to parameter k for scenario i. After calculating the weighted score for each judge–scenario pairing, the average score for each scenario was calculated across all judges as follows:

$$\bar{S_i} = \frac{1}{N}\sum_{j=1}^{N} S_{ij} \quad (2)$$

In this equation, $\bar{S_i}$ is the mean weighted score for scenario i, and N represents the total number of judges involved in the evaluation. This two-step process incorporates both the relative importance of the evaluation criteria and the aggregated judgment of the expert panel into the final scenario metrics used for comparative analysis.

2.5. Component Counts and Workload Reduction

Without merging, the total number of ratings was calculated as follows:

$$Total\ Rating = 5\ parameters \times 6\ scenarios\ per\ phases \times 5\ phases \times 9\ participants\ per\ arm \times 2\ arms = 2700$$

After merging the answers, the total number of ratings was reduced to:

$$Total\ Rating = 5 \times 6 \times 5 \times 1\ pmerged\ answer\ per\ arm \times 2 = 300$$

This reduction preserved the parameter-level information contained in the data while making the evaluation process feasible for the HR experts. The approach also retained the comparative structure of the experiment and avoided unnecessary duplication of substantially similar responses. Therefore, the reduction from 2700 to 300 ratings improved operational efficiency without sacrificing the intended analytical detail or comparative reliability.

2.6. Data Aggregation

For each PMBOK process group, the mean weighted score across all six scenarios was calculated separately for the human and AI arms. These phase-level scores were then used in the comparative analysis to identify trends of AI superiority, human superiority, or equivalent performance. The aggregation procedure maintained the distinction between the five project phases while allowing the overall performance of each participant type to be summarized in a consistent quantitative form. The resulting phase-level measures formed the basis for subsequent statistical testing and cross-phase interpretation.

2.7. Bias Mitigation Controls in Scoring

The scoring process incorporated several controls to minimize evaluator bias. Under blind attribution, evaluators were not informed whether a response had been generated by an AI model or written by a human Project manager. Arm-neutral numbering was used by labeling answers generically, such as Response A and Response B, thereby reducing the possibility of source recognition. The presentation order of human and AI answers was randomized to prevent sequence bias. This scoring design integrates expert-defined relevance weights, remains compatible with PMBOK competency expectations, and reflects the operational realities of downstream oil and gas Project Management in Iran.

The approach also follows the double-combination scoring logic documented in the thesis appendices. The raw table contains 2×1350=2700 scores, calculated as 2 arms × 5 parameters × 6 scenarios × 9 participants × five phases. The merged table contains 2×150=300 scores, calculated as 2 arms × 5 parameters × 6 scenarios × 1 merged answer × five phases. This methodological structure enables the conclusions to be traced directly to a quantified and reproducible evaluation process.

2.8. Data Analysis Strategies

The analysis strategy translates the structured evaluation data collected from the human and AI arms under identical conditions into comparative performance measures for each PMBOK process group and knowledge area. The analysis preserves statistical rigor while maintaining contextual validity for Iran's downstream oil and gas sector. It includes data preparation, weighted aggregation, performance-ratio calculation, phase-level comparison, paired significance testing, effect-size estimation, multiple-testing correction, and reliability and factorability assessment.

2.9. Data Structure from Evaluation Stage

The evaluation framework generates a dataset in which each record contains a Scenario ID ranging from 1–30, representing five phases × 6 scenarios per phase. Each record also identifies the participant type, consisting of either a merged human answer or a merged AI answer, the Parameter ID from 1–5, the parameter weight on an integer scale of 1–5, the raw evaluator score on a Likert scale of 1–11, and the corresponding weighted score calculated by multiplying the raw score by the parameter weight. Under the merged structure, each HR evaluator provides 300 ratings, calculated as 5 parameters × 6 scenarios × 5 phases × 2 arms × 1 merged answer per arm. Without merging, the number would be 2,700 ratings per evaluator, equivalent to 2×1350=27002 \times 1350=27002×1350=2700 scores.

2.10. Cleaning and Preparation

Before statistical analysis, the dataset undergoes consistency checks to identify missing, invalid, or inconsistent scores. Parameter weights are applied to the raw evaluation scores, and the resulting weighted values are normalized across evaluators by converting them into phase-relative percentages in order to control for evaluator stringency or leniency. The data are stored in two forms, namely the raw scoring table with 2×1350 observations and the merged scoring table with 2×150 observations, consistent with the Appendix Tables and the combined-solution scores. These procedures ensure that the subsequent comparisons are based on complete, standardized, and traceable data.

To facilitate direct comparison of performance metrics such as time efficiency, error rate, and quality score, a performance ratio was established between the AI system and the human participants. This ratio normalizes the results and indicates relative superiority across the project phases and evaluation parameters. The ratio is defined as:

$$Ratio_{(P\phi)} = \frac{\bar{X}_{AI(P\phi)}}{\bar{X}_{Human(P\phi)}} \tag{3}$$

Here, $Ratio_{(P\phi)}$ represents the comparison between AI and human performance for parameter p in project phase ϕ. The term $\bar{X}_{AI(P\phi)}$ denotes the mean AI performance for the specified parameter and phase, while $\bar{X}_{Human(P\phi)}$ denotes the corresponding human mean. Values greater than 1 indicate AI outperformance, values below 1 indicate human superiority, and a value equal to 1 indicates equivalent aggregated performance across scenarios.

2.11. Aggregation Across Scenarios

For each PMBOK phase p and participant type t, the phase mean score was calculated as:

$$PhaseMeanScore_{(pt)} = \frac{1}{6}\sum_{s=1}^{6}\left(\frac{\sum_{i=1}^{5} WeightedScore_{(si)}}{\sum_{i=1}^{5} Weighted_i}\right) \tag{4}$$

This procedure produces 10 mean scores, representing five project phases for each of the two participant types. The phase-level measures serve as the baseline for cross-phase comparison and provide a standardized basis for evaluating whether human or AI responses demonstrate stronger performance under equivalent scenario conditions.

2.12. Paired t-test

The core inferential analysis examines whether statistically significant differences exist between paired measurements, such as AI performance and human performance on the same scenario and evaluation parameter. Because the two arms respond to identical task instances, the observations are treated as paired rather than independent. The paired t-test is calculated as:

$$t = \frac{\bar{X}_D}{s_D/\sqrt{n}} \tag{5}$$

In this equation, t represents the calculated test statistic, $\bar{X}_D$ is the mean of the paired differences, defined as $\bar{X}_{AI} - \bar{X}_H$, s_D is the standard deviation of the paired differences, and n is the number of paired observations. In the present study, n corresponds to the number of tasks or scenarios evaluated identically for the human and AI arms.

2.13. Cohen's d Effect Size

Statistical significance alone does not indicate the practical magnitude of the observed difference; therefore, Cohen's d was calculated to quantify the effect size. This measure standardizes the difference between the AI and human means and facilitates interpretation regardless of the original measurement scale. Cohen's d was calculated as:

$$d = \frac{\bar{X}_{AI} - \bar{X}_H}{s_{pooled}} \tag{6}$$

The pooled standard deviation was calculated as:

$$s_{pooled} = \sqrt{\frac{(n_{AI}-1)s_{AI}^2 + (n_{Human}-1)s_{Human}^2}{n_{AI} + n_{Human}}} \tag{7}$$

Here, s and s represent the variances of the AI and human performance metrics, respectively, while n and n denote the sample sizes of the two groups. Following Cohen's (1988) conventional interpretation, a d value of 0.2 indicates a small effect, 0.5 represents a medium effect, and 0.8 or higher denotes a large effect.

2.14. Cross-Phase Comparative Analysis

For each project phase p, a paired comparison was conducted between $PhaseMeanScore_{(pAI)}$ and $PhaseMeanScore_{(pHuman)}$. Descriptive statistics, including the mean, median, and standard deviation, were first calculated to summarize the distribution of scores. Sign tests were then used to determine whether one participant arm consistently outperformed the other after applying the parameter weights. In addition, the Wilcoxon signed-rank test was used for non-parametric paired analysis because the original Likert scores are ordinal, although aggregated means are frequently treated as interval-level measures in comparative research.

Because the analysis involves multiple tests across different parameters and project phases, the probability of Type I errors increases. To control the cumulative error rate while preserving reasonable statistical power, the Benjamini–Hochberg procedure was applied to control the False Discovery Rate at $q = 005$. First, all $m\ p - values$ were ordered in ascending form as $p_1 \le p_2 \le \cdots \le p_m$. Second, critical values were calculated as:

$$p_{(i)}^{*} = \frac{i}{m}q$$

Where q is the selected False Discovery Rate level, set at 0.05 in this study. Third, the largest value of i satisfying $p_i \le p_i^*$ was identified, and all null

hypotheses corresponding to $p - values$ up to and including i were rejected. Alternatively, the FDR-adjusted $p - value$ can be expressed as:

$$p_{FDR(i)} = \frac{p_i m}{i} \tag{8}$$

This procedure ensures that the reported significant findings remain controlled for the increased likelihood of false discoveries resulting from multiple comparisons.

2.15. Double-Combination Tables for High-Level Insights

In addition to the raw and merged tables, which reduce the data from 2×1350 to 2×150, the appendix includes a "double combined" scoring structure containing 2×25=50 scores. These 50 scores are calculated as 2 arms × 5 parameters × 1 scenario per phase × five phases. This higher-level aggregation enables rapid identification of extreme differences at the phase level and supports simplified visualization in Chapter 4, including Figure 4 1 and the corresponding bar graphs. It also provides input for the applied web platform, which classifies scenario phases and recommends the dominant performer.

2.16. Cronbach's Alpha and Internal Consistency

Before analyzing the performance metrics, the quality and internal consistency of the data-collection instrument were assessed. Cronbach's Alpha was used to determine how consistently the items or scale components measured the same underlying construct. The coefficient was calculated as:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} \sigma_i^2}{\sigma_T^2}\right) \tag{9}$$

Here, k represents the number of test items or scale components, σ_i^2 denotes the variance of item i, and σ_T^2 represents the variance of the total score across all items. In social science research, values of $\alpha \geq 070$ are conventionally considered evidence of acceptable internal reliability.

2.17. Kaiser–Meyer–Olkin Measure

The Kaiser–Meyer–Olkin measure was calculated to determine whether the data were suitable for factor analysis. The KMO statistic compares the magnitude of observed correlations with the magnitude of partial correlations and therefore assesses the proportion of shared variance among the variables. It was expressed as:

$$KMO = \frac{\sum r_{ij}^2}{\sum r_{ij}^2 + \sum p_{ij}^2} \tag{10}$$

In this equation, r_{ij} represents the simple correlation coefficient between items i and j, while p_{ij} represents the partial correlation coefficient between the same items after controlling for the remaining variables. Values closer to 1.0 indicate greater suitability for factor analysis, and a value of KMO > 0.60 is generally considered adequate.

2.18. Bartlett's Test of Sphericity

Bartlett's Test of Sphericity was used to determine whether the correlation matrix differed significantly from an identity matrix. The test assesses whether the observed variables are sufficiently interrelated to justify factor analysis and other dimension-reduction techniques. The test statistic was calculated as:

$$x^2 = -\left(n - 1 - \frac{2p+5}{6}\right)\ln|R| \tag{11}$$

Here, n is the sample size, p is the number of observed variables, and $|R|$ is the determinant of the correlation matrix. A statistically significant result, defined as $p < 005$, indicates that the variables share sufficient common variance and that the dataset is appropriate for factor analysis.

2.19. Data Preparation and Experimental Context

This chapter presents the quantitative analysis of the collected data across three analytical stages: Stage 3, which calculates ratios by phase and parameter; Stage 4, which conducts significance testing using the REAL dataset; and Stage 5, which performs paired t-tests after FDR correction. The dataset was generated from responses evaluated by 11 professional Project Management recruiters who assessed human and AI outputs across 30 weighted scenarios. The scenarios were equally distributed among the five project-life-cycle phases, namely (1) Initiation, (2) Planning, (3) Execution, (4) Monitoring and Control, and (5) Closure.

As shown in Table 3 each evaluator assigned scores according to five weighted parameters. These parameters were selected because they represent the most important characteristics of high-quality Project Management decisions in Iran's downstream oil and gas sector. Human and AI responses were exposed to exactly the same scoring framework for every scenario, thereby ensuring comparability. Nine AI reasoning models generated the automated responses, and these outputs were matched against nine human participants with equivalent professional backgrounds.

Table 3 . High-Ranked Parameter Chosen for Assessing Responses

Code	Evaluation Parameter	Description
P1	Effectiveness and Efficiency	Objective performance of the proposed solution.
P2	Acceptability	The extent to which the solution would be accepted by upper management.
P3	Cost and Time Saving	The resource economy represented by the decision.
P4	Accuracy	Technical soundness and internal consistency.
P5	Creativity	Originality or innovation of the proposed approach.

For each scenario, both human and AI responses were evaluated through the same five-parameter scoring framework. All subsequent analyses conducted in Stages 3–5 were derived directly from the weighted means without applying additional transformations. This arrangement preserves the traceability of the original judgments and ensures that the reported comparisons reflect the predefined evaluation structure.

2.20. Bias Mitigation and Ethical Considerations

The research incorporates multiple methodological safeguards to maintain rigor and minimize bias arising from participant awareness, evaluator expectations, or scenario design. Human Project managers were provided only with the essential scenario details and clear response expectations. They were not informed of the comparative nature of the study or of the performance of their G-AI counterparts, thereby reducing the possibility of competitive anxiety, defensiveness, strategic responding, or overconfidence. This procedural blinding supports a more natural representation of professional decision-making.

The HR evaluators were fully blinded to the source of each response and were not informed whether a submission had been produced by a human Project manager or a G-AI model. This double-blind procedure reduces both conscious and unconscious preferences and strengthens the objectivity of the evaluation process. All scenarios and responses were de-identified, and evaluators received standardized scoring rubrics based on the validated industry parameters. The presentation order of human and AI responses was also randomized to reduce sequence-related bias.

Before deployment, all scenarios were reviewed by an independent panel of domain experts to confirm their realism, neutrality, and relevance to downstream Project Management practice. This validation process reduced the possibility that leading questions, ambiguous wording, or unintended contextual advantages would favor either respondent group. The complete study design, including consent procedures, data handling, anonymization, and participant protections, was submitted for ethical review and approval by the institution's ethics board.

Ethical oversight included transparent communication of participant rights, the collection of informed consent, and the implementation of secure data-management procedures. Together, participant blinding, evaluator blinding, expert scenario validation, anonymization, informed consent, and formal ethical review maintain the methodological integrity and ethical standards of the study. These controls strengthen the fairness, robustness, replicability, and authenticity of the comparison between human and AI-driven Project Management strategies in Iran's downstream oil and gas industry.

3. Results and Discussion

This study evaluates AI-generated and human-produced project-management outputs across the five life-cycle phases of Initiation, Planning, Execution, Monitoring, and Closure. Performance was assessed through five parameters, namely Efficiency (P1), Acceptability (P2), Cost/Time Saving (P3), Accuracy (P4), and Creativity (P5). The analytical design integrated Stage 3 ratio analysis, Stage 4 inferential testing, and Stage 5 paired-sample testing. This sequence enabled comparison of relative performance, statistical significance, and practical effect magnitude.

Stage 3 calculated the performance ratio $R = AI/Human$ for each phase–parameter combination, producing 25 cases (5×5). A ratio of 1.00 denotes equivalent performance, values below 1.001.001.00 indicate human superiority, and values above 1.00 indicate AI superiority. Stage 4 applied two-tailed tests at $\alpha = 005$, including Welch's t-test and Mann–Whitney U-test, with Benjamini–Hochberg false discovery rate correction. Stage 5 then employed paired-samples t-tests and Cohen's d to assess scenario-matched practical differences.

The assessment involved 31 independent experts in project management and Iran's downstream oil and gas industry. Their mean professional experience was approximately 20.6 years $(SD \approx 41;\ range = 14 - 30)$, and most held Master's or Doctoral degrees in Industrial Engineering, Management, or Energy Economics. Participants held senior managerial, academic, research, PMO, or consulting roles in organizations such as the Petroleum Industry Research Institute, NIOC, and Farab. They evaluated the outputs using an 11-point Likert scale, strengthening the industry relevance and credibility of the comparison.

Reliability and validity were verified through a two-stage assessment of the Scenario Raw Scores instrument. Initial results showed moderate reliability ($Cronbach's\ \alpha = 072$) and marginally adequate sampling adequacy ($KMO = 0555$); P2 and P5 were highly reliable ($\alpha = 095\ and\ 089$), whereas P1, P3, and P4 required refinement. Following item pruning, overall reliability improved to $\alpha = 076$, while KMO rose to 0.594. Bartlett's test remained significant ($p < 0001$), confirming that the refined data retained suitability for factor analysis and the intended five-factor structure.

The middle 50% of ratios ranged from approximately 0.43 to 1.05, the overall mean ratio was 0.78, and the full range extended from 0.29 to 1.79. Although most ratios clustered between 0.4 and 0.9, a limited upper tail indicated domains of AI strength, particularly structured Monitoring tasks.

Table 4 . Scatterplot Comparison Diagram, Categorized by Parameters.

Phase	Parameter	Mean $R(AI/Human)$	Dominance
Monitoring	P1 – Efficiency	1.79	$AI \gg Human$
Planning	P1 – Efficiency	0.96	$Human > AI$
Execution	P3 – Cost/Time Saving	0.81	$Human > AI$
Closure	P3 – Cost/Time Saving	0.50	$Human \gg AI$
Closure	P2 – Acceptability	0.29	$Human \gg AI$

Table 5 . Phase-wise Comparison

Phase	Significant Comparisons ($p < 0\ 05$)
Initiation	27
Planning	16
Execution	25
Monitoring	24
Closure	22

Parameter-level results revealed that Acceptability (P2) and Creativity (P5) accounted for most significant AI–human differences, whereas Efficiency (P1) and Accuracy (P4) generally showed non-significant differences. Across Phase × Parameter displayed extensive clusters of $p < 001$ for P2 and P5, particularly in Execution and Closure. In contrast, P1 and P4 were predominantly within the neutral or non-significant range.

Table 6 . Parameter-wise Comparison

Parameter	Significance Pattern	Statistical Behaviour	Dominant Performer
P1 – Efficiency	Low	Generally $p > 005$	Parity, $AI \approx Human$
P2 – Acceptability	High	Frequently $p <$ 0001	Human
P3 – Cost/Time Saving	Medium	Significant mainly in Closure	Human
P4 – Accuracy	Low	Generally $p > 005$	Neutral
P5 – Creativity	Very High	Frequently $p < 000$	Human

Stage 5 introduced a stricter within-scenario comparison by pairing AI and human outputs in each of the 30 identical scenarios. After FDR correction, significant differences remained for P2 and P5 in Execution and Closure, and for P3 in Closure. The relevant ttt-statistics ranged from −6.24 to −9.27, while adjusted probability values were either $p < 0001$ or $p = 0003$. All other phase–parameter pairs had FDR-adjusted $p > 005$, indicating that the strongest and most robust differences were selective rather than universal.

Table 7 . Effect Size Estimates for Significant AI–Human Comparisons

Phase	Parameter	t-Statistic	Adjusted $p - Value$, FDR	Direction
Execution	P2 – Acceptability	-8.79	< 0.001	$Human > AI$
Execution	P5 – Creativity	-9.14	< 0.001	$Human > AI$
Closure	P2 – Acceptability	-9.27	< 0.001	$Human > AI$
Closure	P3 – Cost/Time Saving	-6.24	0.003	$Human > AI$
Closure	P5 – Creativity	-9.22	< 0.001	$Human > AI$

Effect sizes confirm that the detected differences were practically meaningful rather than merely statistically significant. Cohen's ddd values were 1.08 for Execution–P2, 1.11for Execution–P5, 1.15 for Closure–P2, 0.89 for Closure–P3, and 1.12 for Closure–P5, corresponding to moderate-to-large or large effects. Across all methods, Human > AI occurred in more than 90% of valid comparisons, and no FDR-adjusted result showed AI superiority. Thus, the integrated evidence from Stages 3–5 indicates AI parity in Efficiency and Accuracy but sustained human superiority in Acceptability, Creativity, and context-sensitive decision-making.

Table 8 . Integrated Summary of Analytical Stages and Comparative Performance Findings

Stage	Analytical Focus	Principal Result	Final Implication
Stage 3	Ratio analysis, $AI/Human$	Mean ratios <1.00 in nearly all combinations	Descriptive human advantage
Stage 4	Independent significance tests	Strong differences in P2 and P5; limited P1/P4 differences	Parameter-specific human superiority
Stage 5	Paired ttt-tests with FDR	P2 and P5 confirmed; P3 significant in Closure	Robust within-scenario validation

5. Conclusions

This study empirically compared generative AI chatbots and human project managers across the five PMBOK process groups in Iran's downstream oil and gas sector. AI achieved equivalent or superior performance in structured, data-intensive activities within Initiation, Planning, and Monitoring and Control, with an average performance difference of only Δ≈0.2 points in the latter phases. Its advantage was particularly evident in analytical consistency, speed, documentation, quantitative risk assessment, and predictive support. However, the results demonstrate that operational efficiency alone cannot represent comprehensive project-management competence.

Human project managers performed significantly better in Execution and Closure, where stakeholder engagement, negotiation, ethical judgment, creativity, and interpretation of organizational context are critical. The observed differences were statistically significant (p<0.05), while stakeholder communication and conflict-resolution tasks favored humans by a Mean Difference of +1.4. Moreover, the correlation between the AI–human performance gap and task predictability (r=+0.68, p<0.01) confirms that AI effectiveness increases as tasks become more structured and rule-based. Accordingly, human expertise remains indispensable when projects involve ambiguity, culturally embedded expectations, and emotionally sensitive decisions.

The findings support a hybrid project-management architecture in which AI functions as a scalable decision-support and augmentation mechanism rather than a substitute for managerial leadership. The proposed framework allocates analytics, monitoring, forecasting, and repetitive documentation to AI, while assigning contextual prioritization, stakeholder persuasion, ethical arbitration, and creative problem solving to human managers. The longitudinal comparison also showed that chatbot performance evolved from versions 7 to 9, highlighting the necessity of documenting model versions and timestamps in future AI evaluation studies. An ancillary experiment involving 250 prompts further found 84.8% accuracy for "very rude" prompts versus 80.8% for "very polite" prompts, indicating that direct wording may sometimes improve machine interpretability by reducing linguistic complexity.

Despite its contributions, the study is limited by its focus on text-based generative models, the specific context of Iran's downstream oil and gas industry, and the temporal scope of chatbot versions 7–9. Future research should therefore develop adaptive multi-criteria performance metrics, incorporate inter-rater reliability indicators such as Cohen's k, and validate the framework across construction, IT, and manufacturing sectors. Implementing the proposed decision-support platform through tools such as n8n, Telegram bots, and industrial WebApp dashboards can translate the findings into real-time managerial practice. Overall, intelligent project management should be understood as a model of structured complementarity in which artificial precision and human insight continuously reinforce one another.

Author Contributions

All authors contributed equally to the conception, preparation, and writing of this manuscript.

Conflict of Interest

M.A. Hatefi, in the capacity of Editor-in-Chief, had no involvement in the selection of reviewers or the management of the peer-review process, and the other authors, while observing publication ethics in referencing, declare the absence of any conflict of interest.

Acknowledgement

All authors gratefully acknowledge financial support from Petroleum University of Technology (PUT).

AI Used Statment

The authors retain full responsibility for the accuracy, originality, integrity, legality, and ethical compliance of all text, data, tables, figures, analyses, references, and conclusions, and have verified all AI-assisted outputs. Any fabrication, falsification, manipulation, plagiarism, copyright infringement, or other misuse of AI remains solely the responsibility of the authors and does not transfer or diminish their accountability for the published work.

References

[1] T. Omarov, G. Taikulakova, *European Project Management Journal*, 2023, **13(2)**, 26–40. https://doi.org/10.56889/bhgn3378

[2] A. Waqar, I. Othman, N. Shafiq, M. S. Mansoor, *Artificial Intelligence Review*, 2023, **56(11)**, 12771–12798. https://doi.org/10.1007/s10462-023-10467-7

[3] E. G. Ochieng, D. Ominde, T. Zuofa, *Technology in Society*, 2024, **79**, 1027101-14. https://doi.org/10.1016/j.techsoc.2024.102710

[4] P. Tsiglianu, N. Romasheva, A. Nenko, *Resources*, 2023, **12(6)**, 2-27. https://doi.org/10.3390/resources12060064

[5] A. Sircar, K. Yadav, K. Rayavarapu, N. Bist, H. Oza, *Petroleum Research*, 2021, **6(4)**, 379–391. https://doi.org/10.1016/j.ptlrs.2021.05.009

[6] K. M. Hanga, Y. Kovalchuk, *Computer Science Review*, 2019, **34**, 1001911-17. https://doi.org/10.1016/j.cosrev.2019.08.002

[7] Elsevier, *Generative AI Policies for Journals*. (2025). https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals.

[8] J. O. Osei, A. Karimu, M. K. Kally, M. Goanue, *The Oil and Gas Industry in Emerging and Developing Economies*. In: The Economics of the Oil and Gas Industry. London: Routledge, 2023, pp. 15–29.

[9] R. Behboud, A. Zarei, A. Azar, S. A. Ebrahimi, *Energy*, 2023, **272**, 1270911-10. https://doi.org/10.1016/j.energy.2023.127091

[10] M. R. Shokouhi, A. Khademvatani, F. Beiky, *Energy Strategy Reviews*, 2024, **52**, 1013481-10. https://doi.org/10.1016/j.esr.2024.101348

[11] H. K. Alfares, *Introduction to Petroleum and Petrochemical Industries. In: Applied Optimization in the Petroleum Industry*. Springer, 2023, pp. 1–23.

[12] M. A. I. Alyatama, *Application of Agile Project Management in Kuwait Oil and Gas Capital Projects*. Doctor of Business Administration, University of Northampton, 2021.

[13] G. Cahya, D. Ramadhan, *Bridging the Time Gap: Exploring Escalation Method for Accurate, Precise, Reliable, and Valid Project Budgeting*. (2023). https://www.mckinsey.com/industries/oil-and-gas/our-insights.

[14] A. Rahman, A. Rahim, *Project Management in Oil and Gas Industry: A Review*. (2017). https://www.researchgate.net/publication/321224310.

[15] M. Rostami, M. A. Ahangari, *Iranian Journal of Oil & Gas Science and Technology*, 2022, **11(4)**, 1–10. https://doi.org/10.22050/ijogst.2023.381640.1665

[16] M. Burhanudin, D. V. Alpha Suandi, *14th Annual International Conference on Industrial Engineering and Operations Management*, Dubai, UAE. 2024.

[17] Project Management Institute (PMI), *PMP Project Management Professional Examination Content Outline*. Newtown Square: PMI Publications, (2021). https://www.pmi.org/-/media/pmi/documents/public/pdf/certifications/new-pmp-examination-content-outline-2026.pdf?rev=b618cf45573e4276a54151e7636c97bf

[18] G. Cairns, G. *Wright, Building on the Basic Method: Power, Ethics and Critical Scenario Method*. In: Scenario Thinking. Cham: Springer International Publishing, 2018, pp. 85–103.

[19] A. Rawat, S. Gupta, T. J. Rao, *Risk Mitigation of CNG Refueling Network for the Eastern Corridor States of India*. Research Report, (2021). https://dr.ddn.upes.ac.in/bitstreams/da11bf4b-3f1e-4ab8-aef7-80db3931d2f5/download.

[20] Z. Aiman Bin, C. Rosli, *Project Risk Management in Malaysian Oil and Gas Industry: Implementation by EPC Local Contractor*. Technical Report, (2020). https://knova.um.edu.my/student_works_2020s/276/.

[21] A. Retmasari, *An Implementation of Business Process Improvement Project in Executing, Monitoring, and Controlling and Closing Phase at PT.XYZ Indonesia*. Case Study Report, (2010).

[22] M. I. Hashfi, T. Raharjo, *International Journal of Advanced Computer Science and Applications*, 2023, **14(9)**, 366–376. https://doi.org/10.14569/IJACSA.2023.0140940

[23] M. Aoyama, S. Yamaguchi, Y. Yamaguchi, *Journal of Japan Society of Sports Industry*, 2019, **29(1)**, 25–37. https://doi.org/10.5997/sposun.29.29_25

[24] M. Hinostroza, P. Chavez, V. Nuñez, C. Raymundo, *Application of PMBOK to Improve the Deadline of Projects in SMEs Engineering Consultancies*. In: Proceedings of the 4th Brazilian Technology Symposium (BTSym'18). 2019, 140, pp. 487–494.

[25] A. Shazia, *Study of Effective Virtual Project Management Through Emotional Intelligence: Empowerment and Leadership Style in Global Projects*. Research Thesis, 2010. http://142.54.178.187:9060/xmlui/handle/123456789/2494.

[26] N. Lebedieva, Innovative Leadership and Corporate Culture as Tools of Effective Management. State and Regions: Economics and Business, 2024.

[27] A. Ahsun, B. Elly, *Optimizing Resource Allocation for Enhanced Project Efficiency*. (2024). https://www.researchgate.net/publication/385560767.

[28] O. Kehinde, International Journal of Science and Engineering Applications, 2025, **14(2)**, 1-17.

[29] F. Fouzia, *Operational Management of Project (COP): Speciality Training Manual. Algeria: Ministry of Higher Education*, 2023.

[30] M. A. Rahaman, F. Z. Rozony, M. S. A. Mazumder, Md. N. Haque, *Academic Journal on Science, Technology, Engineering & Mathematics Education*, 2024, **4(3)**, 44–62. https://doi.org/10.69593/ajsteme.v4i03.88

[31] U. Zaman, S. Nawaz, R. D. Nadeem, *Journal of Open Innovation: Technology, Market, and Complexity*, 2020, **6(4)**, 1681-19. 6(4), 1–19. https://doi.org/10.3390/joitmc6040168

[32] M. A. Rahaman, F. Z. Rozony, M. S. A. Mazumder, Md. N. Haque, *Academic Journal on Science, Technology, Engineering & Mathematics Education*, 2024, **4(3)**, 44–62. https://doi.org/10.69593/ajsteme.v4i03.88

[33] A. Narang, *Journal of Informatics Education and Research,* 2025, **5(2)**, 68-85.

[34] O. Famoti et al., *International Journal of Multidisciplinary Research and Growth Evaluation*, 2025, **6(1)**, 1831–1840. https://doi.org/10.54660/.IJMRGE.2025.6.1-1831-1840

[35] D. Hulett, *Practical schedule risk analysis*. CRC Press, 2016, pp. 1-240. https://doi.org/10.4324/9781315601885

[36] M. M. I. Shamim, *Global Mainstream Journal*, 2024, **1(1)**, 1–6. https://doi.org/10.62304/ijmisds.v1i1.107

[37] C. A. Ezeigweneme, A. A. Umoh, V. I. Ilojianya, A. O. Adegbite, *World Journal of Advanced Research and Reviews*, 2023, **20(3)**, 1713–1730. https://doi.org/10.30574/wjarr.2023.20.3.2707

[38] M. A. Ali, Q. Rafique, *International Journal of Project Management*, 2024, **6(2)**, 26–50. https://doi.org/10.47672/ijpm.2004

[39] B. Monteiro, D. B. R. Monteiro, Supporting Decision Making with Strategic Foresight. *OECD*, 2023, **63**, 1-71. https://doi.org/10.1787/1d78c791-en

[40] J. W. Herrmann, *Engineering Decision Making and Risk Management.* Wiley, 2015, pp. 1–357.

[41] R. Chatterjee, Fundamental Concepts of Artificial Intelligence and Its Applications. 2020, **1(2)**, 13-24. https://www.mathematicaljournal.com

[42] M. Z. Hossain, L. Hasan, M. A. Dewan, N. A. Monira, *International Journal of Management Information Systems and Data Science*, 2024, **1(5)**, 1–18.https://doi.org/10.62304/ijmisds.v1i05.211

[43] G. Kasparov, *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*. London: John Murray, 2017. p.p 1-304.

[44] D. H. Rai, *Tribhuvan University, Institute of Science and Technology*, 2024, 1-24.

[45] P. Fatima et al., *Saudi Journal of Engineering and Technology*, 2024, **9(8)**, 406–418. https://doi.org/10.36348/sjet.2024.v09i08.005

[46] S. Murugesan, *IEEE Intelligent Systems*, 2025, **40(1)**, 20–24.

[47] B. Tekinerdogan, Assessing Artificial Intelligence: A Critique of Pure Computationalism, *Open Science Preprints*, 2024, 1-15.

[48] R. González, *The Chinese Room Revisited: Artificial Intelligence and the Nature of Mind*, Doctor in Philosophy, 2007.

[49] M. A. Monin, V. A. Terekhova, E. V. Ledeneva, *Utopia y Praxis Latinoamericana*, 2020, **25(5)**, 179–186. https://doi.org/10.5281/zenodo.3984228

[50] S. Frei, *Syntax, Semantics and the Room with No View: Why Searle Is Wrong to Think Computers Can't Think*. Doctoral dissertation, Concordia University 1996.

[51] Amazon Web Services (AWS), What Is NLP? Natural Language Processing Explained. (2025). https://aws.amazon.com/what-is/nlp/.

[52] H. Dhamelia, *International Journal for Research in Applied Science and Engineering Technology*, 2023, **11(12)**, 364–372. https://doi.org/10.22214/ijraset.2023.57341

[53] G. M. Biancofiore, D. Di Palma, C. Pomo, F. Narducci, T. Di Noia, *Conversational user interfaces and agents*, In Human-centered AI: An illustrated scientific quest, Cham: Springer Nature Switzerland, 2025, pp. 399-438.

[54] GeeksforGeeks, What Is a Neural Network?. (2025). https://www.geeksforgeeks.org/neural-networks-a-beginners-guide/.

[55] S. Selitskiy, *Kolmogorov's Gate Non-linearity as a Step toward Much Smaller Artificial Neural Networks*. In: International Conference on Enterprise Information Systems, ICEIS Proceedings. Science and Technology Publications, 2022, pp. 492–499.

[56] M. Cilimkovic, *Institute of Technology Blanchardstown, Blanchardstown Road North Dublin*, **15(1)**, 18-30. https://dl.icdst.org/pdfs/files/c9ea365e15acbb738862b8114769e96b.pdf

[57] A. Voichenko, A. Vokrug, *Artificial Intelligence as an Agent of Cognition: Transition from Human Capabilities to Superhuman Modalities and Emerging Ethical Issues*, Anton Voichenko, 2024, PP. 1-15.

[58] M. T. Hicks, J. Humphries, J. Slater, Ethics and Information Technology, 2024, **26(2)**, 2-10. https://doi.org/10.1007/s10676-024-09775-5

[59] IBM, What Is GPT (Generative Pre-trained Transformer)?. (2025). https://www.ibm.com/think/topics/gpt.

[60] OpenAI, GPT-4. (2025). https://openai.com/index/gpt-4/.

[61] IBM, What Is a Chatbot?. (2025). https://www.ibm.com/think/topics/chatbots.

[62] M. R. Varzaneh, S. A. Alibrahim, O. Dakkak, K. M. Karaoğlan, *Conversational Agent Chatbot Classification and Design Techniques*. 2025, pp. 257–292

[63] S. Ghosh, S. Ness, S. Salunkhe, *Journal of Engineering Research and Reports*, 2024, **26(6)**, 327–345. https://doi.org/10.9734/jerr/2024/v26i61184

[64] O. Dobariya, A. Kumar, Mind Your Tone: Investigating How Prompt Politeness Affects LLM Accuracy, *arXiv preprint arXiv*, 2025, 2510.049501-5.

[65] J. Wei et al., *36th International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates, 2022.
[66] M. Vogel, *The Genera*, 2024.
[67] S. A. Haider, S. Tehseen, K. Y. Koay, P. Poulova, B. Afsar, *Acta Psychologica*, 2024, **247**, 1043051-13. https://doi.org/10.1016/j.actpsy.2024.104305
[68] L. Hallo, T. Nguyen, *Administrative Sciences*, 2022, **12(1)**, 2-25. https://doi.org/10.3390/admsci12010004
[69] G. O. Daramola, O. A. Ajala, *Engineering Science & Technology Journal*, 2024, **5(4)**, 1473-1486. DOI: 10.51594/estj/v5i4.1078
[70] S. Salimimoghadam, A. N. Ghanbaripour, R. J. Tumpa, A. K. Rahimi, M. Golmoradi, S. Rashidian, M. Skitmore, *Buildings*, 2025, **15(7)**, 1130-32. https://doi.org/10.3390/buildings15071130
[71] M. Vasudevan, B. Senthil Kumar, *Evolving Strategies for Organizational Management and Performance Evaluation*. 2025, pp. 177–208.
[72] P. Ahrweiler, Participatory Artificial Intelligence in Public Social Services. 2023.
[73] U. Sarfraz, *Journal of Xidian University*, 2024, **18(9)**, 461-480.
[74] L. Emma, *The ethical implications of artificial intelligence: a deep dive into bias, fairness, and transparency.* Retrieved from Emma, L.(2024). The Ethical Implications of Artificial Intelligence: A Deep Dive into Bias, Fairness, and Transparency, 2024, pp. 117-207.
[75] Y. W. Chen, *Plurality in Artificial Intelligence Ethics: Through Collaborative and Democratic Approaches, Degree of International Master in Security*, University of Glasgow, Dublin City University, Charles University, 2024.
[76] H. H. H. Aldboush, M. Ferdous, *International Journal of Financial Studies*, 2023, **11(3)**, 90-108. https://doi.org/10.3390/ijfs11030090
[77] G. B. Mensah, *Africa Journal for Regulatory Affairs (AJFRA)*, 2025, **2024(1)**, 32–45. https://doi.org/10.62839/AJFRA/2024.v1.I1.32-45
[78] G. Dhilipkumar, *International Journal of All Research Education and Scientific Methods (IJARESM)*, 2025, **9(4)**, 1687-1688. https://www.ijaresm.com/ijaresm-journal
[79] A. Rahimian et al., *Engineering, Construction and Architectural Management*, 2025, **3(10)**, 7032-7061. https://doi.org/10.1108/ECAM-09-2023-0958
[80] P. Kordian, *Exploring the Potential of AI to Support Team Effectiveness in Interdisciplinary Product Development*. Delft University of Technology, 2023. PP. 1-132.
[81] A. Rehan, D. Thorpe, A. Heravi, *International Journal of Construction Education and Research*, 2024, **21(2)**, 164–188. https://doi.org/10.1080/15578771.2024.2359378
[82] D. Cawthorpe, *Qeios*, 2023, 1-26.
[83] L. K. Gupta et al., *European Chemical Bulletin*, 2023, **12**, 1126–1135.
[84] U. Shajil, A. Sujith, B. G. Nair, *Comparison of Human Intelligence with Artificial Intelligence*. In: Artificial Intelligence and Biological Sciences. 2025, pp. 68–92.
[85] C. Challoumis, K. Challoumis, *XIX international scientific conference*. London. Great Britain, 2024.
[86] F. A. Atobatele, C. A. Okonkwo, *Journal of Frontiers in Multidisciplinary Research*, 2024, **5(1)**, 60–69. https://doi.org/10.54660/.IJFMR.2024.5.1.60-69
[87] G. Huzooree, M. Yadav, N. J. Dewasiri, *AI and Emotional Intelligence in Project Management: Managing Hybrid Teams in a Tech-Driven Environment*. 2023, pp. 111–130.
[88] K. A. Maspul, M. Ardhin, H. Prasetyo, *Research of Economics and Business*, **1 (1)**, 2025, 43-65. https://doi.org/10.70895/roe.v1i1.31
[89] R. Darwish, D. Cadorin, S. Nair, Decision Making Biases in Project Portfolio Selection and Prioritization: An Exploratory Study of the Rationale Behind Decision Making Leading to Project Portfolio Problems. 2014.
[90] M. A. Ghorbani, AI Tools to Support Design Activities and Innovation Processes. 2023.
[91] A. Jooyafar, M. Shakhsi-Niaei, *19th National Student Conference of Industrial Engineering*. 2021.
[92] V. Stingl, A. Gilchrist, A. Lawani, R. Flin, O. Zwikael, I*nternational Journal of Project Management*, 2025, **43(1)**, 1026781-7. https://doi.org/10.1016/j.ijproman.2025.102678
[93] A. Wiewiora, G. Murphy, *Knowledge Management Research & Practice*, 2015, **13(1)**, 17–30. https://doi.org/10.1057/kmrp.2013.26
[94] M. H. Jarrahi, D. Askay, A. Eshraghi, P. Smith, *Business Horizons*, 2023, **66(1)**, 87–99. https://doi.org/10.1016/j.bushor.2022.03.002
[95] J. Rowley, *Management Research Review*, 2014, **37(3)**, 308–330. https://doi.org/10.1108/MRR-02-2013-0027
[96] M. Revilla, C. Ochoa, *International Journal of Market Research*, 2017, **59(5)**, 557–565. https://doi.org/10.2501/IJMR-2017-039
[97] A. Bryman, *Social Research Methods*. Oxford: Oxford University Press, 2016, pp. 1-432.
[98] S. Brinkmann, S. Kvale, *Doing Interviews*. London: SAGE Publications, 2018, pp. 1–208.
[99] B. DiCicco-Bloom, B. F. Crabtree, *Medical Education*, 2006, **40(4)**, 314–321. https://doi.org/10.1111/j.1365-2929.2006.02418.x
[100] R. E. Levitt, J. C. Kunz, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 1987, **1(1)**, 3–24. https://doi.org/10.1017/S0890060400000111
[101] T. V. Fridgeirsson, H. T. Ingason, H. I. Jonasson, H. Jonsdottir, *Sustainability*, 2021, **13(4)**, 23451-20. https://doi.org/10.3390/su13042345
[102] D. D. Lewis, *Artificial Intelligence and Law*, 2010, **18(4)**, 481–486. https://doi.org/10.1007/s10506-010-9101-0
[103] H. K. Dam, T. Tran, J. Grundy, A. Ghose, Y. Kamei, *41st International Conference on Software Engineering: New Ideas and Emerging Results*. IEEE, 2019, pp. 41–44.
[104] G. Auth, O. Jokisch, C. Dürk, *Online Journal of Applied Knowledge Management*, 2019, **7(1)**, 27–39. https://doi.org/10.36965/OJAKM.2019.7(1)27-39
[105] R. K. Maheshwari, *Journal of Information Systems Engineering and Management*, 2025, **9(4)**, 2543–2558. https://doi.org/10.52783/jisem.v9i4s.12956
[106] A. Barcaui, A. Monat, *Project Leadership and Society*, 2023, **4**, 1001011-12. https://doi.org/10.1016/j.plas.2023.100101
[107] A. Doostmohammadi, *The 5th National Conference on New Achievements in Electrical, Computer and Industrial Engineering*. 2023.
[108] E. Eaton, *AI Magazine*, 2017, **38(2)**, 13–21. https://doi.org/10.1609/aimag.v38i2.2730
[109] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman et al., B. McGrew, *arXiv preprint arXiv*, 2023, **1**, 087741-100. https://doi.org/10.48550/arXiv.2303.08774
[110] S. A. Razavi, A. K. Manshad, L. Kashani, *Energy Strategy Reviews*, 2026, **65**, 102226-30. https://doi.org/10.1016/j.esr.2026.102226
[111] S. A. Razavi, H. Motavali, *Journal of Industrial Engineering and Management Studies*, 2023, **10(2)**, 106-115. https://doi.org/10.22116/jiems.2023.399150.1516
[112] H. Mofid, M. R. Khosravi-Nikou, S. A. Razavi, M. Mohammadian, *Applied Thermal Engineering*, 2024, **245**, 122810-17. https://doi.org/10.1016/j.applthermaleng.2024.122810