RESEARCH ARTICLE

DECIS: An LLM-Based Value Assessment Framework for Targets

Zixuan Zhuang[1], **Haoxiang Cheng**[1], **Changjun Fan**[1] **and Chao Chen**[1,*]

[1] Laboratory for Big Data and Decision, College of Systems Engineering, National University of Defense Technology, Changsha 410073, China

Abstract

Target value assessment is a critical component of operational decision-making; however, existing methodologies suffer from the rigidity of static models in dynamic battlefield environments and the scarcity of specialized domain-specific data. To address these problems, we propose DECIS, a framework for adaptive decision support based on Retrieval-Augmented Generation (RAG). DECIS operates through a cohesive three-stage architecture centered on two core functions: first, RAG is employed to intelligently match optimal assessment strategies to dynamic contexts, thereby mitigating the challenge of limited military corpora; second, the Large Language Model (LLM) component autonomously generates and refines target lists, ensuring decision recommendations are characterized by high temporal validity and situational awareness. Experimental validation confirms that DECIS reduces the Largest Connected Component to 0.6, compared to 0.75 for PageRank and 0.8 for Betweenness Centrality, providing robust decision support and demonstrating competitive performance against established baselines in dynamic target value assessment.

Keywords: target value assessment, retrieval-augmented generation, large language models.

1 Introduction

Modern warfare is undergoing rapid evolution toward intelligent, multi-domain operations where strategic success is fundamentally predicated on the efficient execution of the Observe-Orient-Decide-Act (OODA) loop. This paradigm necessitates real-time assessment of targets, a capability strained by the increasing complexity of multi-domain operations and the expansion of engagement domains [1].

Traditional expert-based assessment systems are increasingly inadequate. They struggle with three core limitations: the inability to systematically extract tactical rules from unstructured data, poor adaptation to dynamic battlefield conditions, and failure to meet stringent real-time processing requirements.

To address these gaps, we propose DECIS, a Dynamic Evaluation and Contextual Intelligence Support framework. Its core innovation leverages a RAG architecture for accurate rule extraction from unstructured text and dynamic knowledge maintenance. A secondary innovation is a traceability

Academic Editor:
Muhammad Attique Khan

Submitted: 05 February 2026
Revised: 19 June 2026
Accepted: 26 August 2026
Published: 21 September 2026

Vol. 3, **No.** 3, 2026.
10.62762/TIS.2026.310611

***Corresponding author:**
Chao Chen
chenc1997@nudt.edu.cn

Citation
Zhuang, Z., Cheng, H., Fan, C., & Chen, C. (2026). DECIS: An LLM-Based Value Assessment Framework for Targets. *ICCK Transactions on Intelligent Systematics*, 3(3), 162–171.

mechanism linking decisions to source data, ensuring the interpretability and auditability vital for reliable decision support.

This paper is organized as follows. Section 2 reviews relevant work. The DECIS framework is detailed in Section 3. Section 4 presents the experimental verification. Finally, Section 5 summarizes the findings and discusses directions for future research.

2 Related Work

2.1 Target Value Assessment Methods

Target value assessment is a fundamental multi-criteria decision-making problem for military resource prioritization, evolving from simple metrics to complex frameworks alongside intelligent warfare.

Early approaches relied on structured expert judgment. The Analytic Hierarchy Process (AHP) decomposes decisions hierarchically using pairwise comparisons to determine factor weights. However, AHP's reliance on static expert knowledge produces context-fixed weights that cannot adapt to dynamic battlefield changes, leading to suboptimal decisions when situations deviate from initial assumptions [2].

To address battlefield uncertainty, researchers integrated AHP with complementary techniques, drawing on Grey Relational Analysis methods for evaluating alternatives under incomplete or fuzzy information [3], and constructed task-specific frameworks for applications such as weapon-target assignment [4] and air combat assessment [5]. Yet these hybrid models introduce scenario-binding rigidity: their indicator systems are designed for specific contexts and cannot transfer to different engagements without extensive manual redesign.

Recent work incorporates advanced theories for systemic complexity. The OODA loop models dynamic assessment processes [6], complex network theory evaluates targets by their systemic influence rather than individual attributes [7], and combined subjective-objective weighting balances expert opinion with data-driven insights [8]. Despite conceptual advances, these approaches rely on static battlefield representations with predefined network structures, node importance logic, and weight fusion mechanisms. Adapting them for evolving situations or new missions remains an expert-driven process lacking the agility required for high-tempo operations.

In summary, while existing research has enriched target value assessment from subjective weighting to network-centric analysis, all methodologies share a fundamental limitation: static, predefined frameworks. Whether through fixed AHP weights, task-specific indicators, or immutable network structures, all require labor-intensive reconfiguration for new circumstances—a critical shortcoming for operations demanding real-time adaptability. Consequently, an intelligent framework is needed that not merely calculates values within fixed paradigms but dynamically generates and optimizes assessment logic in evolving contexts. LLMs, with their emergent reasoning, world knowledge, and in-context adaptation capabilities, offer a promising avenue for enabling systems that understand complex contexts and formulate tailored evaluation strategies on-the-fly without manual re-engineering.

2.2 Large Language Model Reasoning

LLMs exhibit strong reasoning but suffer knowledge gaps in specialized domains like military target value assessment. Advanced prompting (e.g., CoT [9], ToT [10], GoT [11]) enhances interpretability and structured reasoning but cannot supply up-to-date domain knowledge. Retrieval-Augmented Generation (RAG) addresses this through a retrieve-and-generate paradigm integrating dynamic retrieval with LLMs [12]. RAG offers three advantages: dynamic knowledge via updatable bases [13]; retrieval-constrained generation anchoring outputs to facts [14]; and enhanced response quality through optimized retrieval configurations tailored to specific knowledge bases [15].

Recent years have seen growing interest in applying RAG to specialized decision-making tasks. Hao et al. [16] constructed a domain-specific LLM for UAV swarm control by combining an element-based knowledge base with a hybrid retrieval and reranking RAG framework, demonstrating substantial gains in accuracy and human alignment over general-purpose models. Siriwardhana et al. [17] proposed RAG-end2end, which jointly trains the retriever and generator to adapt RAG to domain-specific knowledge bases, significantly improving performance in open-domain question answering beyond Wikipedia-scale corpora. Zhang et al. [18] integrated RAG into intelligent networking agents, enabling real-time knowledge retrieval and interactive decision support in dynamic environments. Similarly, Fan et al. [19] provided a systematic survey of RAG-meets-LLM paradigms, cataloguing multi-strategy retrieval architectures including

adaptive retrieval, query rewriting, and text filtering that collectively enhance response quality in knowledge-intensive scenarios.

Despite these advances, existing RAG-based systems applied to specialized domains share common limitations. First, most rely on static knowledge bases, limiting real-time decision support. Second, retrieval strategies lack adaptivity, reducing relevance in evolving scenarios. Third, interpretability and traceability are underemphasized, undermining trust and verification. DECIS addresses these limitations by embedding RAG within a structured decision-support architecture for military target value assessment. By integrating domain-specific retrieval with structured reasoning, DECIS ensures rigorous and actionable assessments.

3 DECIS Framework

DECIS comprises three core components: vector knowledge base establishment, evaluation method selection, and target list generation. Each of these components is elaborated upon sequentially in the subsequent sections, as illustrated in Figure 1.

3.1 Knowledge Base Establishment

Although general-purpose LLMs excel at handling general tasks, they frequently lack sufficient depth in domain-specific knowledge. This shortcoming is particularly acute in the domain of military target value assessment, which is characterized by data scarcity, restricted accessibility, and inherent quality variability. Consequently, these models lack the specialized knowledge requisite for their direct application, particularly concerning established assessment methodologies. To address this gap, the DECIS framework leverages a dual-layer architecture, comprising an alternative methodology base and an expert knowledge base. This architecture is designed to integrate specialized domain knowledge, thereby providing a verifiable theoretical foundation for enhancing foundation models in this domain.

The establishment of the expert knowledge base, illustrated in Figure 1 (A), integrates data from three distinct dimensions: operational doctrine, historical disposition records, and expert experience. Operational doctrine includes publicly accessible performance data for weapons and equipment, alongside standardized decision-making frameworks from governmental or military bodies. Historical disposition records comprise compilations of historical cases from combat operations and military exercises, capturing dynamic tactical patterns and outcome feedback. Expert experience involves tacit knowledge gathered through methodologies such as the Delphi technique and structured interviews. In contrast, the alternative method base is constructed from empirically validated assessment methodologies synthesized from the academic literature.

In military target value assessment research, domain-specific corpora are relatively scarce and exhibit task-coupled characteristics, with original texts often embedded within multi-task contexts such as operational planning and situational analysis. Concurrently, this domain is characterized by high information redundancy and low semantic density. Constructing a knowledge base directly from raw corpora would likely lead to redundant computational resource consumption and dilution of core knowledge representation. To mitigate this, this study introduces an LLM-driven text summarization technique, achieving directed extraction of target value assessment knowledge through structured prompt engineering.

The prompt system adopts a quadruple architecture. First, the background definition clarifies the tactical level and operational scenario. Second, objective constraints prescribe the assessment dimensions and granularity standards for extraction. Third, exemplar illustration dictates the output template and parameter units. Finally, formatting standards supply illustrative examples from typical tactical scenarios.

Storing information directly in textual format presents three principal drawbacks: retrieval reliance on exact keyword matching, poor performance in complex semantic queries, and difficulty in supporting vector similarity computation and high-dimensional data retrieval needs. Consequently, this study constructs the expert knowledge base by converting textual rules from domain experts into vector embeddings stored in a vector database. This knowledge base preserves the semantic content of the text in high-dimensional vector form, makes the relationships explicit, and supports efficient retrieval and reasoning based on semantic similarity, thereby providing contextual knowledge support for target value assessment.

Furthermore, DECIS functionally distinguishes the method repository from the expert knowledge base. The method repository stores candidate assessment methods in a structured textual format, detailing evaluative models, procedures, and applicable conditions. This textual format enables expert

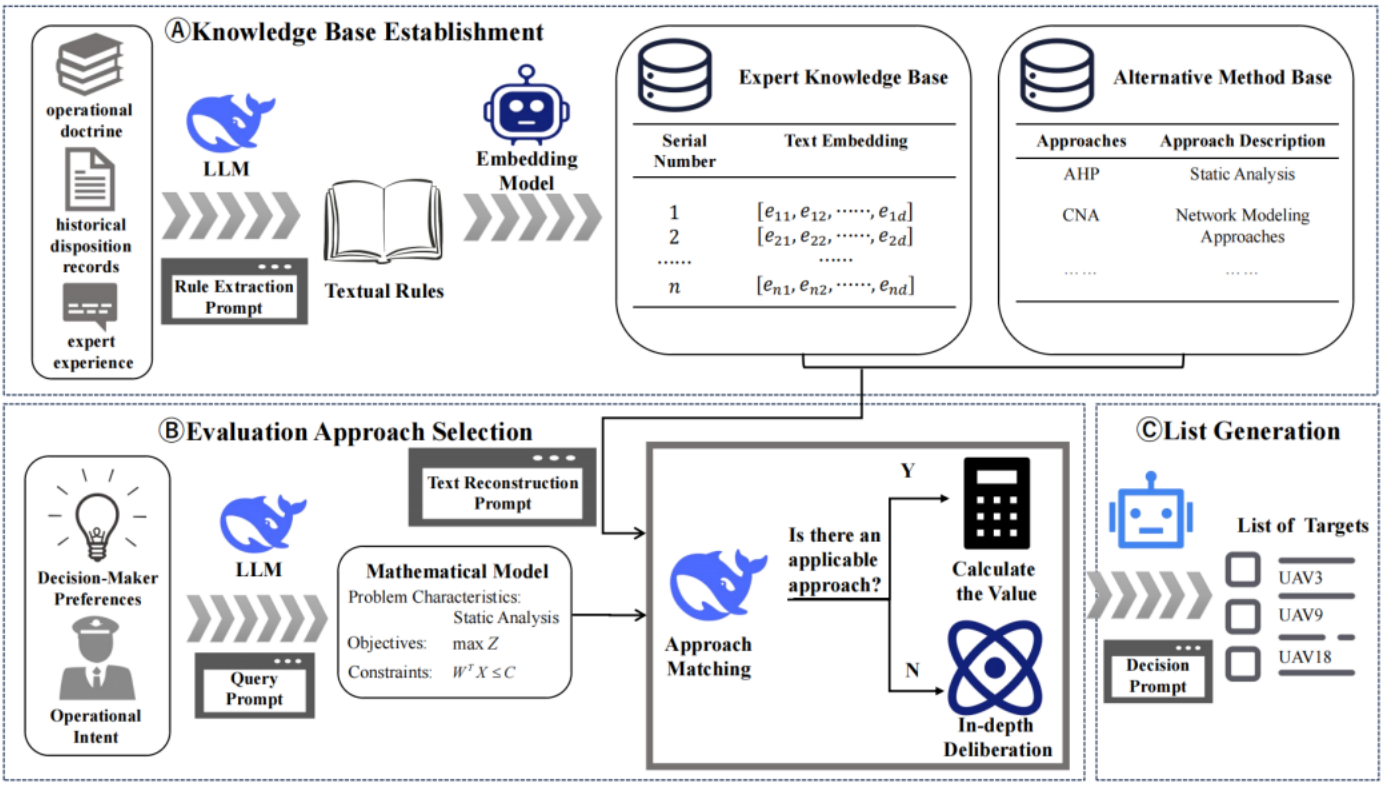

Figure 1. DECIS consists of three principal components. Following the establishment of the expert knowledge base and the alternative method base in stage (A), the large model in stage (B) identifies and matches suitable methods according to the decision maker's input. Subsequently, stage (C) generates the target value and the strike list.

verification and iterative updates, ensuring method interpretability and practical implementability. In contrast, the expert knowledge base stores abstracted knowledge in vectorized form. The expert knowledge base provides a theoretical foundation for method selection; the method repository translates this knowledge into executable assessment pathways. Together, they form a "knowledge-to-execution" chain, supporting the target value assessment workflow from acquisition to implementation.

3.2 Evaluation Approach Selection

Contemporary target value assessment methodologies are well-developed for specific objects and scenarios. However, complex battlefield environments require dynamic adaptation of these approaches to evaluate targets with divergent characteristics. A key limitation is their inability to parse latent commander inputs—such as operational intent, tactical preferences, and target characteristics—to infer appropriate assessment methods. To overcome this limitation, we leverage LLMs to construct a mathematical representation of command intent, thereby facilitating autonomous method selection through similarity computation with pre-established knowledge bases.

The evaluation method matching mechanism, illustrated in Figure 1 (B), requires dual-dimensional input from the commander: operational intent and command preferences. Operational intent characterizes the ultimate objective of the action, possessing global guiding attributes. Command preferences reflect the decision-maker's risk tolerance and prioritization of specific indicator sets, directly influencing weight configuration within the evaluation model and the attribute weighting process subsequent to method selection.

Given the inherent fuzziness and redundancy of natural language, its direct input can easily induce semantic ambiguity, leading to information loss or decision bias. This research employs deep semantic understanding and information extraction techniques based on large language models to transform commander natural language directives into a formatted mathematical model. This model comprises three core elements: problem characteristics, objectives, and constraints. Problem characteristics delineate parameters such as the dynamic/static nature of the decision scenario, hierarchical structural features, and data scale, providing a feature-dimension basis for assessment

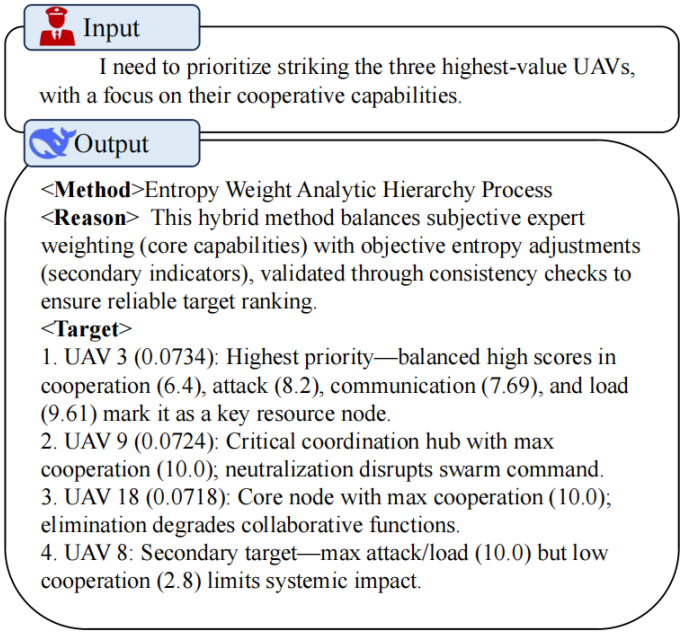

Figure 2. Example of target list output.

method matching. Objectives involve defining decision variables to formalize command intent into objective function expressions. Constraints systematically integrate resource limitations and operational costs within the battlefield environment, constructing a complete mathematical programming framework.

The chunked vectorized storage mechanism employed for the expert knowledge base and the method repository presents challenges during retrieval, including data discretization and semantic ambiguity. This study designs specialized text reconstruction prompts to drive the large language model to integrate discrete texts into coherent contextual segments, discarding ambiguous information and expressing content using clear and concise language, thereby facilitating effective matching between expert methods and the problem model.

During the retrieval and matching process, DECIS may encounter expert-method mismatch problems under complex tactical situations. When candidate methods exhibiting significant similarity to the current problem model are detected, DECIS performs dynamic weight allocation based on command preferences to generate a composite evaluation value. If no suitable method exists, a deep reasoning engine is triggered to produce interpretable decision recommendations. This dual-mode processing mechanism effectively ensures decision continuity across varying matching scenarios.

3.3 Target List Generation

Although large language models possess natural language generation capabilities, the absence of a standardized output framework for military target value assessment results leads to unclear methodological justification and insufficient explanation of solution rationales. To resolve these issues, this study designs the target list generation architecture illustrated in Figure 1 (C), with a representative output shown in Figure 2.

In complex decision-making scenarios, the reasoning processes of LLMs often exhibit prolonged path dependencies, resulting in information redundancy and focal ambiguity within raw outputs. To address this limitation, the present research develops a structured decision prompt template that employs semantic constraint mechanisms to achieve directed extraction of decision elements and supporting evidence, ultimately producing a concise and focused target system value list.

4 Experiments

This study designs an experiment to validate the solving efficacy of DECIS within a specific military decision-making scenario through a full-process instantiation.

4.1 Scenario

This experiment addresses threat assessment for a heterogeneous cluster of twenty unmanned aerial vehicles (UAVs) in an information-enabled operational environment, as depicted in Figure 3. The UAV cluster forms a tactical network with missions spanning reconnaissance, jamming, and precision strikes. The objective is to identify and degrade critical nodes through dynamic target value assessment.

Each UAV is modeled as a network node characterized by four normalized capabilities: Attack, Load, Communication, and Cooperation. Original data are generated via Monte Carlo simulation and normalized to a $[0, 10]$ scale using min–max normalization:

$$a_k^{\text{norm}} = \frac{a_k - \min(A_k)}{\max(A_k) - \min(A_k)} \times 10, \tag{1}$$

where A_k represents the complete set of the k-th attribute. The network topology is derived from node attributes. Node degree d_i is mapped from communication capability c_i^{norm} as:

$$d_i = 2 \times \left\lfloor \frac{c_i^{\text{norm}} - 2}{2} \right\rfloor \tag{2}$$

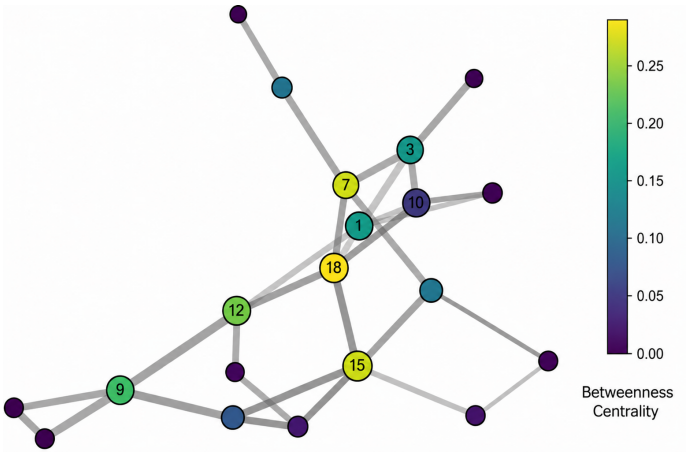

Figure 3. Scenario of the UAV network.

where d_i is set to zero when $c_i^{\text{norm}} < 2$.

Edge weight w_{ij}, representing the cooperative potential between nodes i and j, is calculated as the average of their combined normalized attributes:

$$w_{ij} = \frac{1}{8} \times \sum_{k=1}^{4} \left(a_{ik}^{\text{norm}} + a_{jk}^{\text{norm}} \right). \tag{3}$$

The experimental framework integrates a vector database containing structured assessment methods (Table 1) and expert knowledge (Table 2). For a given query, DECIS uses the BAAI General Embedding (BGE) model [20] to encode the query and compute the cosine similarity against the knowledge base:

$$\text{similarity} = \frac{A \cdot B}{\|A\| \, \|B\|}. \tag{4}$$

The best-matching method is retrieved and applied to compute node values. Finally, a structured decision prompt guides the LLM to output a formatted target strike list with method justification and prioritized recommendations.

4.2 Modeling

The knowledge base establishment phase comprises three sequential steps: data preprocessing, rule extraction, and vectorized storage. Initially, operational regulations from various military branches, equipment operation specifications, records of historical engagement cases, and expert experience documents undergo format standardization. Precise feature extraction is then achieved through rule extraction prompts. Ultimately, embedding vectors are generated and stored in a vector database, thereby forming the expert knowledge base.

Table 1. Alternative methods for value assessment.

Method	Description
Entropy Weight–AHP	A hybrid approach integrating subjective experience with objective data analysis.
Complex Network Theory	A methodology based on network modeling and analysis.
Bayesian Network	Suitable for dynamic analysis and scenarios with limited sample sizes.
PROMETHEE Method	A technique for multiple-attribute decision making.
Grey Relational Analysis	Evaluates based on geometrical curve similarity and supports dynamic analysis.
PageRank Algorithm	A link-analysis algorithm adapted for network importance ranking.
.....	.....

The expert method selection mechanism in DECIS follows a structured pipeline. For a query like "UAV threat level assessment", the framework first transforms the input via a query prompt into a mathematical model that annotates the problem's hierarchical structure (Figure 4). Concurrently, it uses text-processing prompts to contextually reconstruct discrete knowledge base data into a structured, logical representation. A full-library vectorized retrieval then computes spatial correlations using cosine similarity. The DeepSeek-R1 reasoning model [21] serves as the inference engine to score the most similar method, and a final decision prompt outputs a target list containing the selected method, selection rationale, top-three solutions, and their justifications.

4.3 Results Analysis

In response to the query for identifying the three UAVs with the highest strike value, DECIS selected the Entropy Weight–AHP method. This method was chosen for its ability to balance the subjective emphasis on key indicators like attack capability with the objective calibration of other metrics, ensuring a consistent and hierarchical assessment.

As reported in Table 3, the top three recommended targets are UAV 3, 9, and 18. Cooperative capability

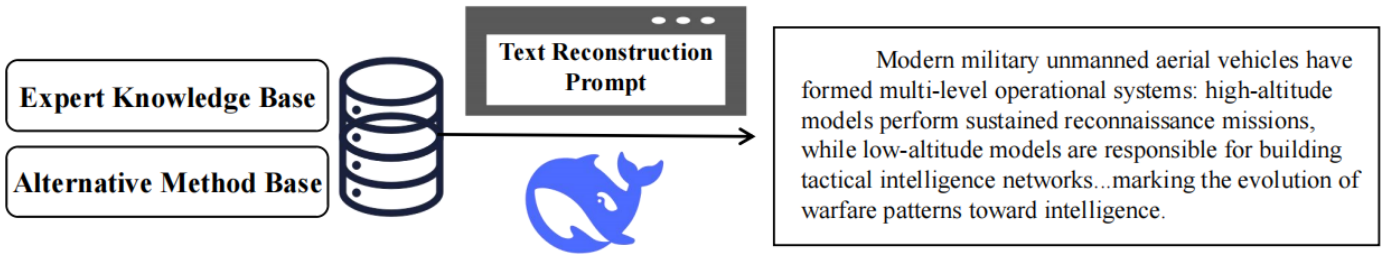

Figure 4. Contextual reconstruction results.

Table 2. Expert knowledge base examples.

Rule ID	Rule Statement
Rule 1	Low-altitude miniaturized rotorcraft achieve information coverage in urban street fighting through terrain-matching algorithms and passive listening modes.
Rule 2	In system-of-systems confrontation, the efficiency of the sensor-shooter closed loop holds greater significance than the performance of individual weapon systems.
.....	.....
Rule n	Compared to traditional reconnaissance aircraft, low-altitude miniaturized unmanned aerial vehicles demonstrate enhanced capabilities in covert infiltration and special operations.

Table 3. UAV value scores.

Rank	ID	Score	Rank	ID	Score
1	3	0.0734	11	7	0.0503
2	9	0.0724	12	20	0.0489
3	18	0.0718	13	4	0.0460
4	8	0.0655	14	5	0.0407
5	19	0.0623	15	17	0.0393
6	11	0.0596	16	14	0.0373
7	15	0.0594	17	16	0.0366
8	12	0.0582	18	2	0.0320
9	1	0.0553	19	6	0.0240
10	10	0.0543	20	13	0.0128

was the dominant factor: UAV 9 and 18 were selected due to their maximum cooperative score (10.0) combined with strong auxiliary capabilities. UAV 3, while having a lower cooperative score (6.4), achieved a high composite value through its prominent attack (8.2) and communication (7.69) capabilities, indicating a systemic advantage. In contrast, UAV 8—despite having maximum attack and load capabilities—ranked only fourth due to its low cooperative capability (2.8), demonstrating the framework's prioritization of network-centric value over individual performance.

A comparative experiment was designed with the size of the Largest Connected Component (LCC) adopted as the core evaluation metric to elucidate the critical role of entropy. The rate of change in LCC size serves as an effective indicator to characterize the degradation of the target system's structural robustness. As shown in Figure 5, which illustrates the results of network disintegration via entropy-AHP and AHP respectively, the disintegration effect of entropy-AHP consistently equals or outperforms that of the standard AHP method with the removal of an identical number of nodes.

Given that the entropy-AHP method primarily incorporates individual target characteristics, a further comparative analysis was designed to demonstrate the systemic value of targets identified by DECIS in comparison to those identified by network-based approaches. For this comparison, three widely utilized network centrality metrics were chosen: Degree Centrality, Betweenness Centrality, and PageRank. The systemic value of targets, as evaluated by these four distinct approaches, was quantified by the sequential removal of nodes in descending order of their assigned scores and by monitoring the proportional reduction in the LCC, as illustrated in Figure 6.

The LCC decay curves for all methods exhibit similar initial trends, indicating a consensus regarding the identification of the most prominent core nodes within the network. However, a significant divergence emerges upon the removal of the first three nodes. The proposed method, together with Degree Centrality, attains a superior LCC ratio of 0.6, in contrast to 0.8 for Betweenness Centrality and 0.75 for PageRank. This indicates that the proposed method is more

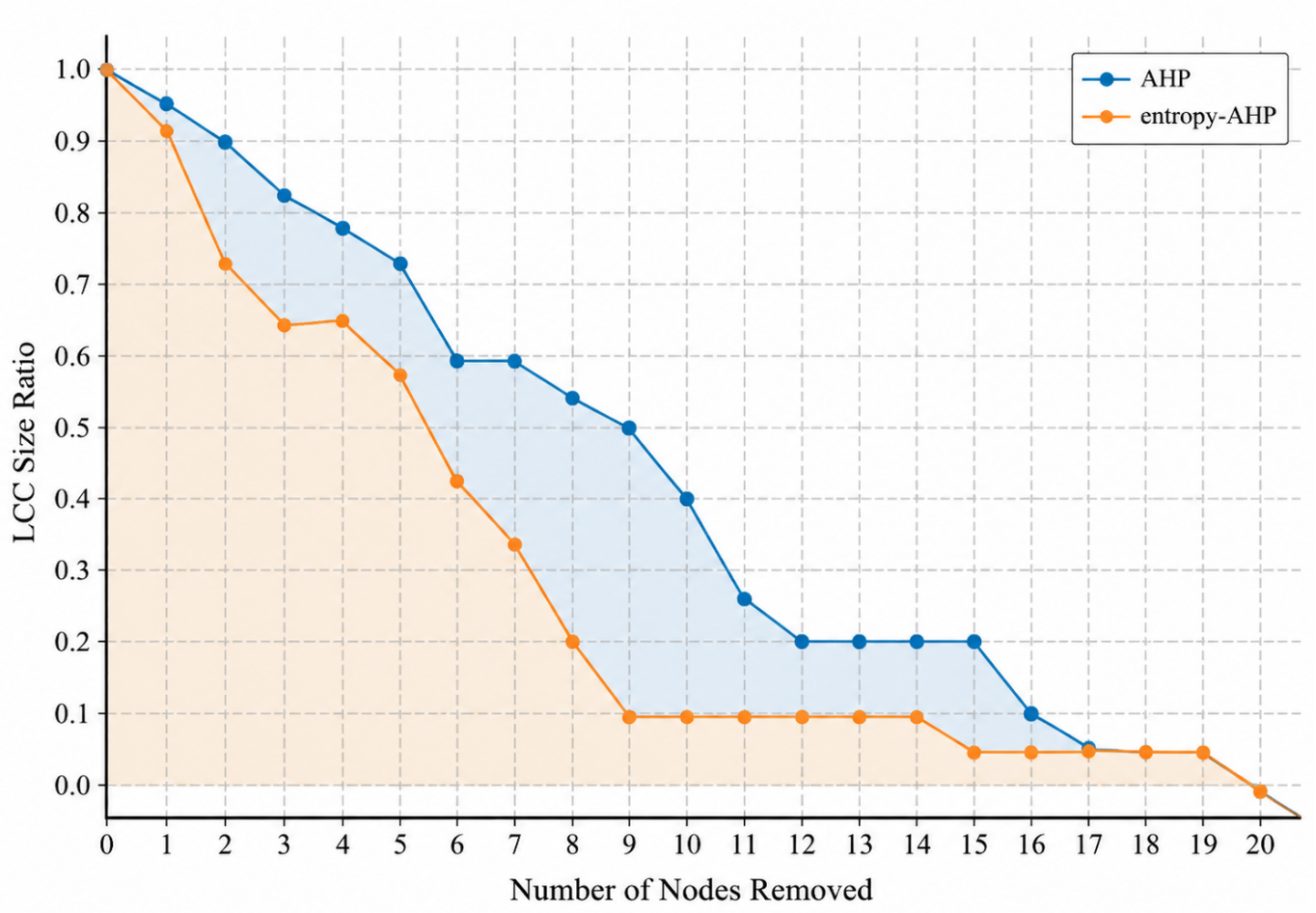

Figure 5. A comparison of network disintegration effectiveness between entropy-AHP and AHP.

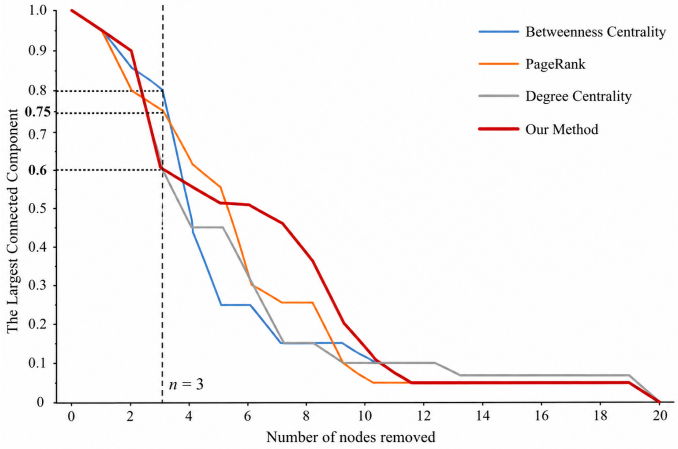

Figure 6. Network collapse results.

effective at reducing network connectivity, as it selects nodes based on their collective impact on the overall system performance rather than their isolated prominence. Although both methods achieve the same LCC ratio at this stage, DECIS derives its node rankings from a semantic assessment of multi-dimensional capabilities—including attack, load, communication, and cooperation—rather than purely topological degree information, thereby offering greater interpretability and adaptability in dynamic operational contexts where graph topology alone is insufficient to characterize target value.

5 Conclusion

This study constructs an intelligent target value assessment framework, designated DECIS, which is based on LLMs. Experimental results validate the efficacy of this framework. DECIS demonstrates a robust capability to leverage expert methods stored within its knowledge base, selecting appropriate analytical techniques based on the characteristics of the target information. Ultimately, it provides decision support to commanders through natural language interpretations of its results.

Future research will focus on enhancing DECIS to process and interpret the ambiguous or implicit intent of commanders. Advancing this capability is expected to further improve the framework's utility in supporting complex value assessment tasks within

dynamic operational environments.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

DeepSeek-R1 was used in two distinct capacities in this work. In its role as a *research tool*, it served as the inference engine of the proposed DECIS framework, responsible for method scoring and target list generation. Separately, in its role as a *writing aid*, it was used to assist with language editing and translation during manuscript preparation. All scientific content, methodology, experimental results, and conclusions were independently developed and verified by the authors, who take full responsibility for the integrity of this work.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Manolache, I. C. (2023). The role of multi-domain operations in modern warfare. *Land Forces Academy Review, 28*(3), 163-170. [CrossRef]

[2] Luo, R., Huang, S., Zhao, Y., & Song, Y. (2021). Threat assessment method of low altitude slow small (LSS) targets based on information entropy and AHP. *Entropy, 23*(10), 1292. [CrossRef]

[3] Qi, Z. F., & Wang, G. S. (2014). Grey synthetic relational analysis method-based effectiveness evaluation of ewcc system with incomplete information. *Applied Mechanics and Materials, 638*, 2409-2412. [CrossRef]

[4] Li, J., Wu, G., & Wang, L. (2024). A comprehensive survey of weapon target assignment problem: Model, algorithm, and application. *Engineering Applications of Artificial Intelligence, 137*, 109212. [CrossRef]

[5] Yin, Y., Zhang, R., & Su, Q. (2023). Threat assessment of aerial targets based on improved GRA-TOPSIS method and three-way decisions. *Math. Biosci. Eng, 20*(7), 13250-13266. [CrossRef]

[6] Lu, X., Ma, H., & Wang, Z. (2022). Analysis of OODA loop based on adversarial for complex game environments. *arXiv preprint arXiv:2203.15502*. [CrossRef]

[7] Wang, Y., Liang, J., & Wang, S. (2022). A New Approach for Target Capability Assessment and Selection in Complex Situations. *Computational Intelligence and Neuroscience, 2022*(1), 3146683. [CrossRef]

[8] Wu, W., Jie, W., Luo, A., Liu, X., & Luo, W. (2025). Data-fusion-based algorithm for assessing threat levels of low-altitude and slow-speed small targets. *Sensors, 25*(17), 5510. [CrossRef]

[9] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems, 35*, 24824-24837.

[10] Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems, 36*, 11809-11822.

[11] Besta, M., Blach, N., Kubicek, A., Gerstenberger, R., Podstawski, M., Gianinazzi, L., ... & Hoefler, T. (2024, March). Graph of thoughts: Solving elaborate problems with large language models. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 38, No. 16, pp. 17682-17690). [CrossRef]

[12] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems, 33*, 9459-9474.

[13] Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*. [CrossRef]

[14] Shuster, K., Poff, S., Chen, M., Kiela, D., & Weston, J. (2021, November). Retrieval augmentation reduces hallucination in conversation. In *Findings of the Association for Computational Linguistics: EMNLP 2021* (pp. 3784-3803). [CrossRef]

[15] Şakar, T., & Emekci, H. (2025). Maximizing RAG efficiency: A comparative analysis of RAG methods. *Natural Language Processing, 31*(1), 1-25. [CrossRef]

[16] Hao, J. X., Chen, L., & Meng, L. (2025). Advancing large language models with enhanced retrieval-augmented generation: Evidence from biological UAV swarm control. *Drones, 9*(5), 361. [CrossRef]

[17] Siriwardhana, S., Weerasekera, R., Wen, E., Kaluarachchi, T., Rana, R., & Nanayakkara, S. (2023). Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering. *Transactions of the Association for Computational Linguistics, 11*, 1-17.

[CrossRef]

[18] Zhang, R., Du, H., Liu, Y., Niyato, D., Kang, J., Sun, S., ... & Poor, H. V. (2024). Interactive AI with retrieval-augmented generation for next generation networking. *IEEE network, 38*(6), 414-424. [CrossRef]

[19] Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., ... & Li, Q. (2024, August). A survey on rag meeting llms: Towards retrieval-augmented large language models. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 6491-6501). [CrossRef]

[20] Xiao, S., Liu, Z., Zhang, P., Muennighoff, N., Lian, D., & Nie, J. Y. (2024, July). C-pack: Packed resources for general chinese embeddings. In *Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval* (pp. 641-649). [CrossRef]

[21] DeepSeek-AI. (2025). DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv preprint arXiv:2501.12948*. [CrossRef]

Zixuan Zhuang is with the Laboratory for Big Data and Decision, College of Systems Engineering, National University of Defense Technology, Changsha, Hunan, China. (Email: zhuangzixuan22@nudt.edu.cn)

Haoxiang Cheng is with the Laboratory for Big Data and Decision, College of Systems Engineering, National University of Defense Technology, Changsha, Hunan, China. (Email: hx_chenggfkd@nudt.edu.cn)

Changjun Fan is with the Laboratory for Big Data and Decision, College of Systems Engineering, National University of Defense Technology, Changsha, Hunan, China. (Email: fanchangjun@nudt.edu.cn)

Chao Chen is with the Laboratory for Big Data and Decision, College of Systems Engineering, National University of Defense Technology, Changsha, Hunan, China. (Email: chenc1997@nudt.edu.cn)