Advances in Human Factors of Transportation, Vol. 224, 2026, 191–200
https://doi.org/10.54941/ahfe1007846

AI Competency Model for Aerospace Engineering Managers

Huizhong Hu[1], Ye Tian[2], Ao Jiang[1], Mingzhao Xu[3], Rui Hou[4], Yong Su[5], Yong Ni[1], Shengchi Li[6], Chenyu Ouyang[7], Jun Zhou[7], Mulan Huang[1], and Jinglin Zhang[1]

[1]Nanjing University of Aeronautics and Astronautics, Nanjing, China
[2]Tsinghua University, Beijing, China
[3]Northwestern Polytechnical University, Xian, China
[4]Supervisory Unit Stationed at the Second Institute of China Aerospace Science and Industry Corporation, Beijing, China
[5]Supervisory Unit Stationed at the China Electronics Institute, Beijing, China
[6]Jiangsu Aerospace Industrial Design Research Institute, Nanjing, China
[7]Nanhua University, Hengyang, China

ABSTRACT

As crewed spaceflight and deep-space exploration accelerate, aerospace engineering organisations are adopting digital and AI-enabled management to meet higher mission cadence, stringent compliance, and evidence-based assurance needs. Yet generic competency frameworks do not adequately capture the role-specific competencies required under safety-critical and highly traceable governance. This study developed and validated an AI competency model and closed-loop assessment framework for aerospace engineering managers. Competency constructs were elicited via RepGrid interviews ($n = 30$) and validated using PCA ($n = 217$) and CFA ($n = 209$). The resulting five-dimensional model showed good internal consistency (Cronbach's $\alpha = 0.900$) and acceptable fit (RMSEA = 0.026; SRMR = 0.045). Expert judgements ($n = 14$) were used to model interdependencies and derive system-aware weights via DEMATEL-DANP. In a recruitment scenario with five candidates (H1-H5) rated by four experts, an improved VIKOR method produced a compromise ranking and gap diagnosis; H5 remained top-ranked across the tested V settings ($V = 0.1$–0.9). The framework provides interpretable ranking and diagnostic outputs to support human-centered, AI-enabled, traceable decisions for selection and development across the project lifecycle.

Keywords: Aerospace engineering management, AI competency, Human-centered aerospace systems, Space sustainability, Repertory grid, Dematel-Danp, Vikor

INTRODUCTION

Aerospace programmes pursuing crewed lunar-Mars exploration and long-duration deep-space missions face increasing governance burden: higher mission cadence, growing data volumes, and stricter evidence-chain requirements demand decisions that remain traceable, verifiable, and accountable (NASA, 2021; NASA Office of Inspector General, 2025). AI is increasingly embedded in engineering management to support planning, requirements translation, configuration control, verification, and

Received February 6, 2026; Revised April 6, 2026; Accepted April 26, 2026; Available online July 20, 2026

risk workflows; however, mis-specified requirements can cause goal drift, overreliance on model outputs can weaken verification, mishandling data and confidentiality boundaries can trigger compliance violations, and inadequate responses to AI failures can introduce new safety hazards. Accordingly, international guidance emphasises human-centered risk management and trustworthy governance for safety-critical domains, offering transferable principles for aerospace contexts (ECSS, 2024; NIST, 2023; ISO/IEC, 2023; EASA, 2023; Jiang et al., 2018, 2022a, 2022b, 2023a, 2023b, 2025).

Existing competency frameworks mainly target generic project management skills, broad digital competencies, or AI usage by technical staff and do not operationalise the role-specific competencies required of aerospace engineering managers under mission assurance and evidence-chain constraints (Lematta, McDermott and Dominguez, 2024; AIAA, 2024).

In light of the aforementioned issues, this study constructs and validates an AI competency assessment framework tailored for aerospace engineering management roles. This framework models and evaluates the artificial intelligence capabilities of aerospace engineering managers within safety-critical and highly regulated project environments. The framework validates a five-dimensional model comprising 20 characteristics, derives interpretable global weightings for impact perception prioritisation, and provides traceable decision support through candidate ranking, gap diagnostics, and robustness checks.

METHODS

(1) Study design.

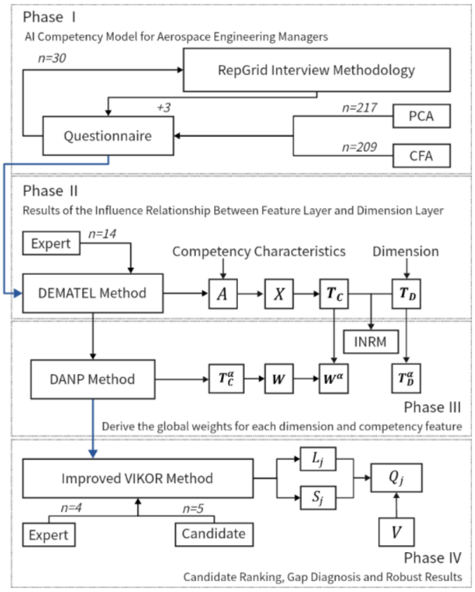

Figure 1: Research framework flowchart.

This study employs a closed-loop, four-phase design comprising competency construct elicitation, structural exploration and validation, influence-network modelling and weight derivation, and candidate ranking with gap diagnosis (Fig. 1).

In Phase I, Repertory Grid (RepGrid) interviews with aerospace engineering managers (n = 30) elicited role-specific AI competency constructs. A two-round questionnaire study then explored and validated the measurement structure: PCA (n = 217) was used for preliminary dimensional classification, followed by CFA (n = 209) to validate the five-factor measurement model and test a second-order structure across dimensions. In Phase II, DEMATEL was applied to expert ratings (n = 14) to derive total influence matrices at both feature and dimension levels and generate an influence network relationship map (INRM). In Phase III, DEMATEL-based ANP (DANP) produced global weights for dimensions and features. In Phase IV, five candidates (H1–H5) were rated by four experts (n = 4), and an enhanced VIKOR method generated compromise ranking and gap diagnosis. Robustness was examined via inter-rater reliability (ICC) and a V-parameter sensitivity analysis (V=0.1, 0.3, 0.5, 0.7, 0.9).

(2) Participants.
Participants were aerospace engineering management personnel with substantial project experience (≥5 years in aerospace projects and ≥3 years in managerial roles), covering core management tasks. Participants were recruited from commercial aerospace companies, launch vehicle enterprises, and related organisations, and had basic exposure to AI-enabled management workflows. Participation was voluntary with informed consent; the protocol was approved by Nanjing University of Aeronautics and Astronautics and all identifiers were anonymised.

(3) Constructing the competency model.
Using RepGrid triadic comparisons, we elicited bipolar constructs from aerospace engineering managers (n = 30) by repeatedly probing "what/why" with concrete examples until construct saturation. Semantically equivalent constructs were consolidated through researcher agreement and adjudication, yielding 42 candidate constructs. A three-channel screening strategy (mention rate, importance rating, and expert deliberation) retained constructs that were either frequently mentioned (≥42%) or considered critical for safety/compliance and accountability. The retained characteristics were operationalised as questionnaire items (7-point Likert) and examined via PCA with Varimax rotation (n = 217; primary loading ≥0.40; loading gap ≥0.15; ≥3 items per factor), followed by CFA in AMOS 29.0 on an independent sample (n = 209) to test both first-order five-factor and second-order structures. Reliability and validity were evaluated using Cronbach's α, CR, AVE, and discriminant validity criteria, with bootstrap estimation used for parameter stability.

(4) Derivation of influence relationship networks and weights.
Fourteen experts independently assessed the direct influence between competency attributes using a 5-point scale. Group consensus was evaluated via the average relative error rate (α), with $\alpha < 5\%$ required before aggregation. The aggregated direct-influence matrix A was then standardised to yield the direct influence matrix X. Series convergence was verified using the spectral-radius criterion $\rho(X) < 1$ to ensure the invertibility of $(I - X)$ (Du and Zhou,

2019). Based on X, the DEMATEL total-relation matrices at the criterion and dimension levels were derived to generate the INRM. Subsequently, DANP transformed the DEMATEL results into an ANP supermatrix; the weighted supermatrix was iterated until convergence, with a threshold of for the difference between consecutive iterations, yielding global weights for each dimension and competency characteristic (Chen and Lin, 2018; Kuo, Hsu and Li, 2015; Falatoonitoosi et al., 2014; Jiang et al., 2022c; Jiang et al., 2020).

(5) Candidate assessment, ranking and gap diagnosis.
To validate applicability in practical selection scenarios, five genuine applicants for aerospace engineering management positions were assessed. Four experts rated each competency characteristic using a 0-10-point scale (0 = barely demonstrated; 10 = outstanding performance). Scores were aggregated by equal weighting, with each candidate's characteristic score computed as the arithmetic mean of expert ratings. Inter-rater reliability was examined using ICC under a two-way random-effects model (ICC(2,1) for single-measure and ICC(2,4) for average-measure), reporting 95% confidence intervals to justify mean aggregation (Shrout and Fleiss, 1979; Koo and Li, 2016).

Using global weights derived from DANP, an enhanced VIKOR procedure computed the normalised gap ratio as $r_{ij} = (10 - g_{ij})/10$, where denotes the mean expert rating for candidate j on characteristic i (with $g_i^*=10$ and $g_i^- = 0$). The weighted total gap (S_j) and maximum weighted gap (L_j) were calculated and combined into Q_j. Sensitivity analysis set V = 0.1, 0.3, 0.5, 0.7, and 0.9 to test ranking robustness. In small-sample internal selection contexts, VIKOR outputs were interpreted as shortlist-supporting compromise solutions and evaluated against acceptable advantage and stability conditions rather than claiming a uniquely dominant winner (Opricovic and Tzeng, 2004).

RESULTS

(1) Construction and validation of a five-dimensional, 20-item AI competency model.

Based on RepGrid-derived constructs and subsequent screening, 20 competency characteristics were retained to represent AI competency requirements for aerospace engineering managers under safety-critical, highly regulated, and evidence-chain governance. In the exploratory sample (n = 217), internal consistency was strong (Cronbach's α = 0.900). PCA supported a five-component structure (Varimax rotation) explaining 60.397% of the total variance, yielding a five-dimensional, 20-item scale.

Based on semantic clustering and governance implications, the five dimensions were labelled as follows: (A) Aerospace AI Use-Case Enablement (scenario expression, team motivation, pilot iteration, requirement translation); (B) AI Compliance and Security Assurance (resilient compliance advancement, confidential red-line controls, AI ethical awareness, data security); (C) AI Risk Control and Management (confidential digital asset governance, mission debriefing AI optimisation, AI usage logging and traceability, AI failure response); (D) AI Cognitive Readiness (foundational

AI understanding, AI learning competency, AI outcome scrutiny, proactive mindset towards AI transformation); and (E) Lifecycle AI Orchestration for Project Management (AI-PM integration, management of human-AI decision boundaries, lifecycle AI use-case identification, full-cycle orchestration and conflict resolution).

In the validation sample (n = 209), CFA supported the proposed structure with acceptable fit for both the first-order model (x^2/df = 1.146, CFI = 0.989, TLI = 0.987, RMSEA = 0.026, SRMR = 0.045) and the second-order model (x^2/df = 1.151, CFI = 0.988, TLI = 0.986, RMSEA = 0.027, SRMR = 0.048). Dimension-level reliability and convergent validity were adequate (Cronbach's α = 0.832-0.854; CR = 0.803-0.839; AVE = 0.506-0.566), and discriminant validity was supported by the Fornell-Larcker criterion (Fornell and Larcker, 1981). A Harman single-factor test suggested limited common method bias (first factor = 42.247%(< 50%)).

(2) Results of the influence relationship between feature layer and dimension layer.

The INRM indicates directional couplings among competency elements rather than independent additive effects (Fig. 2). At the feature level, the most salient leverage points within each dimension were AI requirement translation (a4), confidentiality red-line management (b2), AI failure response (c4), AI outcome scrutiny (d3), and full-cycle AI orchestration and conflict resolution (e4). At the dimension level, Lifecycle AI Orchestration for Project Management (E) showed the highest prominence (k+d = 5.454), followed by AI Risk Control and Management (C; 5.358) and AI Compliance and Security Assurance (B; 5.228). C (0.192), B (0.157), and E (0.127) were net driving dimensions(k–d), whereas Aerospace AI Use-Case Enablement (A; –0.290) and AI Cognitive Readiness (D; –0.190) were primarily influenced dimensions.

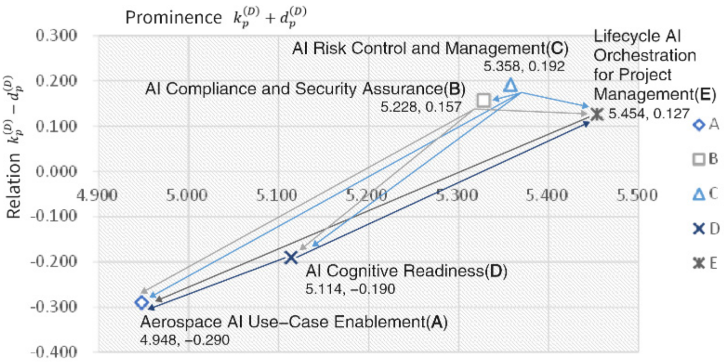

Figure 2: Dimensional INRM showing prominence (k+d) and relation (k–d) for the five dimensions; arrows indicate dominant influence directions.

(3) Candidate ranking, gap diagnosis and robustness results.

Four experts rated five candidates (H1–H5) on 20 characteristics. Inter-rater reliability supported mean aggregation (ICC(2,1) = 0.723, 95% CI [0.648, 0.789]; ICC(2,4) = 0.912, 95% CI [0.881, 0.937]) (Shrout and Fleiss, 1979; Koo and Li, 2016). VIKOR ranking was robust across V = 0.1–0.9, with H5 consistently top-ranked (Opricovic and Tzeng, 2004). However, the observed advantage over the second-best candidate was small ($\Delta Q = 0.004 < 0.25$), indicating limited separability; therefore, results are interpreted as shortlist-oriented decision support rather than a uniquely dominant winner. To support follow-up review, dimension-level gap ratios were visualised as a heatmap (Fig. 3), highlighting where each candidate deviates most from the ideal profile and guiding targeted situational interview probes.

The heatmap results reveal distinct variations in the distribution of candidates' weaknesses. H5 exhibits the smallest overall gap, with the lowest discrepancy in the E dimension (0.19), indicating its comprehensive competency in "end-to-end orchestration and coordination" is closer to the ideal standard. H3 approaches the ideal most closely in the D dimension (gap of 0.20), demonstrating a relative advantage in AI readiness. H4 exhibits substantial gaps across dimensions A/D/E (≥0.53), suggesting its weaknesses are likely manifested as concurrent deficiencies in "use-case implementation, cognitive verification, and full-cycle coordination". Consequently, within scenarios featuring coexisting compromise solutions, heatmaps furnish structured foundations for secondary screening and subsequent situational interviews. Their applied value lies in transforming "ranking" into a closed-loop output encompassing weakness identification, follow-up question design, evidence verification, and remediation pathways. This transition logic from evaluation to action aligns with usability and interpretability requirements in human-centered systems research (Koo and Li, 2016).

	A	B	C	D	E	Q(V=0.5)
H1	0.21	0.32	0.36	0.41	0.21	0.164
H2	0.46	0.22	0.23	0.48	0.44	0.198
H3	0.32	0.32	0.27	0.20	0.33	0.154
H4	0.55	0.31	0.41	0.53	0.53	0.251
H5	0.29	0.31	0.33	0.30	0.19	0.150

0.00 0.40 1.00 Gap ratio (0–1): darker = larger gap

Figure 3: Heatmap of dimension-level gap ratios (0-1) for five candidates (H1–H5), aggregated using DANP global weights; the rightmost column reports Q_j at V = 0.5 (lower is better).

DISCUSSION

This study shows that AI competency in aerospace engineering management is best conceptualised as a governance-embedded, human-centered competency set rather than a collection of generic AI skills. The validated five-dimensional structure indicates that, under safety-critical and highly regulated conditions, managers' AI competence spans not only use-case enablement and cognitive readiness, but also security assurance, risk control, and lifecycle orchestration. Practically, this supports the interpretation that AI empowerment for this role is constrained by traceability, accountability, and evidence-chain integrity: competency becomes meaningful only insofar as it sustains auditable process outputs and decision responsibility after AI integration.

The mechanism evidence further clarifies why this role-specific framing matters. The influence-network results imply that the competency system exhibits directional coupling rather than independent, additive effects. In aerospace engineering governance, where delivery performance depends on cross-disciplinary coordination and closed-loop control, such coupling is expected: weaknesses in requirement translation, confidentiality boundary control, failure response, outcome scrutiny, or full-cycle orchestration can propagate into downstream verification gaps and accountability breaks. Accordingly, the high-prominence "leverage points" identified in the network should be interpreted as training and assessment anchors, not because they are universally important in all industries, but because they plausibly mediate the transition from AI as a local productivity tool to AI as a component embedded in a mission-assurance evidence chain. This provides a human factors explanation for why "AI proficiency" alone is insufficient in this context: the decisive competency is maintaining verifiable, explainable, and compliant work practices when model outputs, human judgement, and organisational responsibility interact.

At the application level, the ranking results are most informative when interpreted as decision support under limited separability, not as a definitive hiring verdict. The stability of the top candidate across preference settings indicates robustness of the overall recommendation direction; however, the lack of acceptable advantage highlights a realistic managerial condition: in high-end shortlists, candidates may be close enough that a single scalar ranking cannot justify high-confidence decisions. This is precisely where a human-centered evaluation framework should add value: by transforming "who ranks higher" into traceable diagnostic evidence indicating where to verify, what to probe in situational interviews, and which gaps to address through development plans. The dimension-level gap profiling operationalises this transition by making shortcomings explicit and reviewable, thereby aligning selection and development with the accountability requirements of aerospace engineering governance.

LIMITATIONS AND FUTURE WORK

This study has several limitations. First, the small candidate pool (n=5) used in the fourth-stage application based on VIKOR is limited in scale. Therefore, this application should be viewed as a scenario-based feasibility validation within a closed-loop evaluation process, rather than proof of universal operational effectiveness. Future work should replicate this application using larger candidate pools and multiple decision scenarios to test its robustness under real-world organisational constraints. Second, although the framework targets aerospace engineering managers requiring high reliability, confidentiality, and compliance, different missions vary in risk levels, process maturity, and evidence chain specifications. Practical deployment necessitates scenario calibration. Future research should develop standardised calibration protocols and report cross-mission measurement invariance when data permits. Model validation in this study primarily relies on internal structural evidence, while standard-related evidence remains insufficient. Subsequent research should incorporate external validity anchors and test predictive or concurrent validity in phased assessment scenarios. Finally, the current assessment partially relies on expert ratings and self-reported importance data, which may be influenced by evaluator bias and information asymmetry. Future research should integrate longitudinal and context-sensitive data (time, task type, performance metrics) and introduce periodic reassessments. This would transform the framework from a one-time scoring system into a continuous, evidence-based, and traceable governance system for aerospace project management talent.

CONCLUSION

This study constructed and validated a five-dimensional, 20-item AI competency assessment framework tailored for aerospace engineering management positions. Exploratory and confirmatory results support its stable measurement structure and robust reliability and validity. Furthermore, the findings indicate that the five dimensions can be subsumed under higher-order general factors, enabling hierarchical interpretation. Using DEMATEL-INRM for influence relationship identification and DANP for weight derivation, the study reveals directional coupling structures among competency elements and identifies key competencies with greater leverage in engineering closed-loop implementation. By enhancing candidate ranking and sensitivity testing through an improved VIKOR approach, this study validated the model's applicability in real-world selection scenarios. Under conditions of "acceptable advantage deficiency," it achieved decision outputs more aligned with engineering management practice by utilising a set of compromise solutions and a dimensional gap heatmap. This enables evaluation results to directly support secondary screening, interview follow-ups, and gap-oriented development plan formulation. Overall, this framework not only answers "Which candidate should be prioritised?" but also addresses "What are the grounds for prioritisation? Where are the shortcomings? How can they be strengthened?" It thus provides traceable, actionable support for role-candidate matching and competency development within the context of intelligent transformation in aerospace engineering management.

ACKNOWLEDGMENT

We thank the teams from Nanjing University of Aeronautics and Astronautics, Tsinghua University, and University of South China for their support during the study preparation. This work was also supported by the National Social Science Fund of China [No. 23FYSB049]; National Natural Science Foundation of China [No. 52405276; No. 52375269].

REFERENCES

AIAA (2024) Safety, trust, and ethics considerations for human-AI teaming in aerospace control. In: AIAA SciTech 2024 Forum.

Chen, S.-H. and Lin, W.-T. (2018) “Analyzing determinants for promoting emerging technology through intermediaries by using a DANP-based MCDA framework”, Technological Forecasting and Social Change, 131, pp. 94–110.

Du, Y.-W. and Zhou, W. (2019) “New improved DEMATEL method based on both subjective experience and objective data”, Engineering Applications of Artificial Intelligence, 83, pp. 57–71.

EASA (2023) EASA artificial intelligence roadmap 2.0: A human-centric approach to AI in aviation. European Union Aviation Safety Agency.

ECSS (2024) ECSS-E-HB-40-02A: Space engineering-Machine learning handbook. European Cooperation for Space Standardization.

Falatoonitoosi, E., Leman, Z., Sorooshian, S. and Salimi, M. (2014) “Expanded DEMATEL for determining cause and effect group in bidirectional relations”, The Scientific World Journal, 2014, Article ID 103846.

Fornell, C. and Larcker, D.F. (1981) “Evaluating structural equation models with unobservable variables and measurement error”, Journal of Marketing Research, 18(1), pp. 39–50.

ISO/IEC (2023) ISO/IEC 23894:2023 Information technology-Artificial intelligence—Guidance on risk management. International Organization for Standardization / International Electrotechnical Commission.

Jiang A, Fang W, Liu J, et al. (2023) The effect of colour environments on visual tracking and visual strain during short-term simulation of three gravity states. Applied Ergonomics 110: 103994.

Jiang A, Foing BH, Schlacht IL, et al. (2022) Colour schemes to reduce stress response in the hygiene area of a space station: a Delphi study. Applied Ergonomics 98: 103573.

Jiang A, Foing BH, Yao X, Gong Y, Li G, Yang H, Westland S (2025) Twelve-Day Space Station Isolation Simulation: The Effect of Virtual Nature Therapy on Depression in an Isolated, Enclosed Environment. Space: Science & Technology 5: 0361.

Jiang A, Gong Y, Yao X, et al. (2023) Short-term virtual reality simulation of the effects of space station colour and microgravity and lunar gravity on cognitive task performance and emotion. Building and Environment 227: 109789.

Jiang A, Schlacht IL, Yao X, et al. (2022) Space habitat astronautics: multicolour lighting psychology in a 7-day simulated habitat. Space: Science & Technology 2022: 9782706.

Jiang A, Yao X, Schlacht IL, et al. (2020) Habitability study on space station colour design. In: Proceedings of the AHFE 2020 International Conference. Springer, pp. 507–514.

Jiang A, Yao X, Westland S, et al. (2022) The effect of correlated colour temperature on physiological, emotional and subjective satisfaction in the hygiene area of a space station. IJERPH 19(15): 9090.

Jiang A, Yao X, Zhou J (2018) Research on path planning of real-time obstacle avoidance of mechanical arm based on genetic algorithm. The Journal of Engineering 2018(16): 1579–1586.

Koo, T.K. and Li, M.Y. (2016) “A guideline of selecting and reporting intraclass correlation coefficients for reliability research”, Journal of Chiropractic Medicine, 15(2), pp. 155–163.

Kuo, T.-C., Hsu, C.-W. and Li, J.-Y. (2015) "Developing a green supplier selection model by using the DANP with VIKOR", Sustainability, 7(2), pp. 1661–1689.
Lematta, G.J., McDermott, P.L. and Dominguez, C. (2024) "Practical guidance for human-machine teaming assurance of AI-enabled systems". In: Proceedings of SPIE—Assurance and Security for AI-enabled Systems, 13054, 130540D.
NASA (2021) NPR 7120.5F: NASA space flight program and project management requirements. NASA Online Directives Information System (NODIS).
NASA Office of Inspector General (2025) Revitalizing the Deep Space Network to support NASA's growing space exploration program. 9 April.
NIST (2023) Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology.
Opricovic, S. and Tzeng, G.-H. (2004) "Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS", European Journal of Operational Research, 156(2), pp. 445–455.
Shrout, P.E. and Fleiss, J.L. (1979) "Intraclass correlations: Uses in assessing rater reliability", Psychological Bulletin, 86(2), pp. 420–428.