Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
April 24, 2025
A Regional Conference of the Society for Industrial and Systems Engineering

An Earned Value-Centered Assessment of Project Health and Artificial Intelligence Intervention Efficacy for the Department of Veterans Affairs

Harrison Brownlee, Gavin Butler, Griffin Lamb, Samuel Lee, Joseph Valchar, James Schreiner, and Matthew Wolfe

Department of Systems Engineering, United States Military Academy, West Point, New York 10996

Corresponding author's email: samuelplee@outlook.com

Author Note: The authors would like to thank the Department of Systems Engineering at the United States Military Academy and the United States Department of Veterans Affairs (VA) Office of Construction and Facilities Management (CFM). In CFM, special thanks to Mr. Ross Davidson, Mr. Glenn Elliott, Mr. Michael Crane, Mr. Dan Gall, and the CFM team for providing data on Risk Assessments (CSRA) of VA projects. The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

Abstract: The Department of Veterans Affairs, with the U.S. Army Corps of Engineers, is developing a method to monitor real-time project health and assess AI's potential for early intervention in "at-risk" projects. This research introduces the Project Risk Integrated System Metrics model, which evaluates ongoing projects using integrated risk metrics. The model utilizes earned value management and other indicators to develop a deterministic, additive risk model that evaluates cost, schedule, and scope. Data sources include Cost-Schedule-Risk Assessments and the Corps of Engineers' Financial Management System. To evaluate AI's utility in risk mitigation, the team used the Project Management Professional exam as a proxy, assessing ChatGPT 4o on two hundred questions across key project domains. The study evaluated response accuracy and explanation quality. The paper will present how the model delivers value to project managers and illustrates that AI performs well in recommending interventions when a project's health is in jeopardy.

Keywords: Project Risk Integrated System Metrics (PRISM), Artificial Intelligence (AI)

1. Introduction

Assessing project health and identifying risks early is critical to support adherence to budget, schedule, and scope in construction projects. Based on meetings with key leaders at CFM, the review found that the organization currently relies on CSRAs, risk registers, and unstructured after-action reviews (AAR) to assess project performance and name risks. This study improves the processes by introducing an additive risk model that integrates traditional earned value management metrics with key project performance indicators to enhance risk assessment. The model processes these metrics to generate a Project Risk Integrated System Metrics (PRISM) score, a dynamic risk rating that reflects project health at any given phase or milestone. The team validated the PRISM model by comparing its output with historical data from three completed VA medical facility projects. This evaluation ensures the model's accuracy in predicting project risk trends. To further enhance risk assessment, this study employs ChatGPT 4o, an advanced AI tool, to analyze risk model outputs and provide targeted intervention recommendations. The team confirmed its effectiveness by assessing AI against Project Management Professional (PMP) exam questions, systematically categorizing them into key project management domains—human resources, cost, schedule, and scope—to ensure alignment with industry best practices. By integrating AI-driven analysis, the model quantifies project risk and suggests mitigation strategies, enabling enhanced decision-making and risk management for VA medical facility projects.

1.1 Literature Review

The literature review examines cost management and risk assessment frameworks used by CFM. Maintaining project health requires a combination of strict cost baselines and effective risk mitigation strategies. Establishing a cost baseline ensures project budgets encompass all estimated costs for resources, materials, and labor while adhering to a clear, actionable roadmap that minimizes delays and cost overruns (Fahad, 2024). In the context of the VA, the Cost Estimating Manual aligns with best practices by emphasizing the integration of cost, time, and scope, using structured methodologies for cost estimation, and iterating estimates as projects evolve. The manual highlights the importance of independent, certified professionals and QA/QC measures to ensure consistency and credibility in cost management (Cost Estimating Requirements for Veterans Affairs

Facilities, 2022). Furthermore, addressing gaps in communication and project understanding between the VA and USACE, and adopting tools like DD Form 1391 for scope definition and improved financial reporting could reduce project risks and streamline processes (USACE, 2024). Expanding upon the CSRA process, which gathers valuable risk data, the development of a tool to show recurring risk factors could help address overall risk in future projects. By using advanced tools and improving collaboration, the VA can enhance infrastructure management, reduce costs, and deliver higher-quality care to veterans.

Figure 1 illustrates the Organizational Responsibility Structure and 'Fit' of the PRISM model with AI intervention feedback loop for USACE and CFM. These agencies manage projects exceeding $100 million, ensuring alignment with congressional mandates. The ER 5-1-11 Project Delivery Business Process (PDBP) establishes a structured framework that enables USACE to manage projects efficiently (USACE Business Process, 2018). By defining milestones and phases at each level, ER 5-1-11 helps effective communication and mitigates risk (USACE Business Process, 2018). Figure 1 also depicts the hierarchical decision-making structure across tactical, operational, and strategic authority levels. It shows how PRISM and Critical Commanders Information Requirements (CCIR) data move through these levels, allowing AI-driven interventions and resource adjustments to support informed decision-making. The strategic authority level integrates the Joint Project Management Office (PMO), ensuring project management consistency across participating organizations. This regulatory framework supports project timelines and strengthens project health through structured governance.

This regulatory framework supports project timelines and strengthens project health through structured governance. Building on this foundation, the current study extends a 2024 Department of Systems Engineering capstone project that introduced an additive value model to calculate project risk, using it as a basis to develop a more comprehensive methodology for modeling and assessing project health (Batt, 2024).

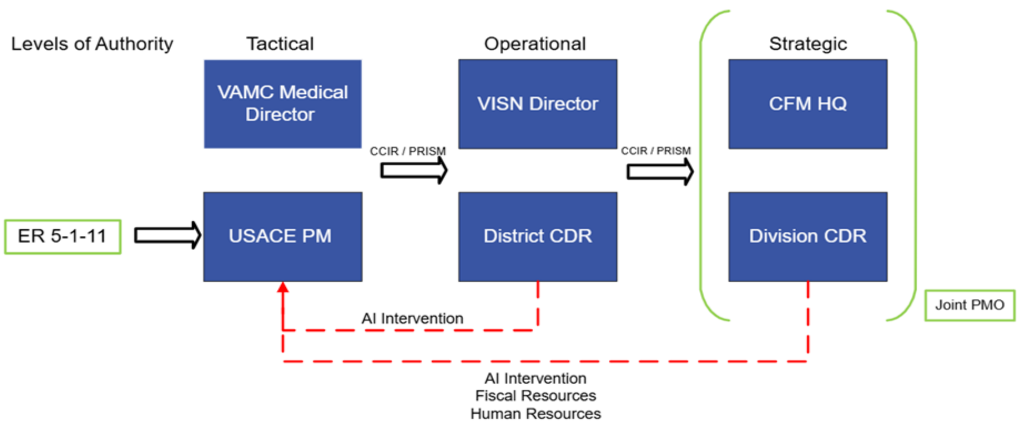

Figure 1: Organizational Responsibility Structure and 'Fit' of the PRISM model with AI Intervention Feedback Loops

2. Methodology

2.1 Systems Thinking and Value Modeling

The Systems Decision Process guides this modeling effort through an additive value modeling approach as seen in Equation 1 that includes value functions, an additive value model, and weighting to assign a final numerical value to a project (Driscoll, Parnell, & Henderson, 2011). This analysis names seven key parameters that, based on client interactions, research, and project management ability, contribute most to risk in CFM construction projects. The value model represents project risk and incorporates these parameters into a risk model.

$$v(x) = \sum_{i=1}^{n} w_i \, v_i \, (x_i) \qquad (1)$$

In this equation, $v_i\,(x_i)$ represents the raw input value for each risk metric, and w_i denotes the weight assigned to that metric, reflecting its relative importance. Equation 1 defines the additive value function used to quantify the PRISM model.

Figure 2 outlines the PRISM Model, a hierarchical framework for evaluating risk in VA construction projects. The model's objective is to predict overall project risk by aggregating key performance data across four subgoals: tracking schedule and cost data, analyzing project changes, assessing contractor quality, and incorporating CSRA data. Each subgoal is supported by quantifiable metrics labeled as "More Is Better" (MIB) or "Less Is Better" (LIB), guiding the desired direction of

ISBN: 97819384969-9-8

performance. Notably, the model includes the Cost Performance Index (CPI), which compares actual cost to planned budget, and the Schedule Performance Index (SPI), which measures project progress compared to planned schedule. The team synthesizes these, along with other metrics such as contingency usage, value engineering adjustments, and contractor quality scores, into a single PRISM score. This integrated, additive approach enhances risk visibility and supports more informed decision making.

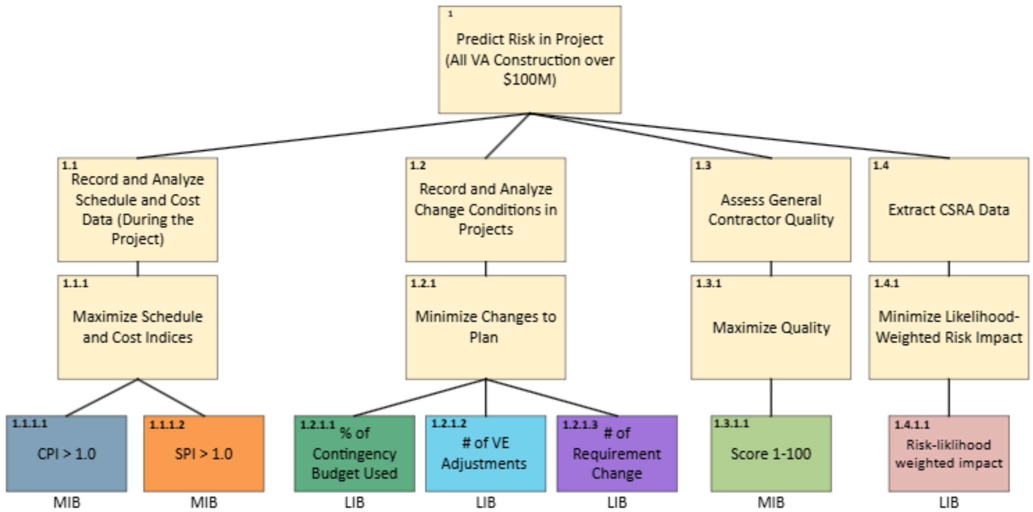

Figure 2: PRISM Model

2.2 Novel Modeling Approach

As seen in Figure 3, the methodology quantifies project risk in a clear and accessible manner, enabling the client to take actionable steps to improve project health. The process starts by modeling risk deterministically. Then, it incorporates uncertainties into the calculations and presents the results in an easily digestible and shareable format. Finally, a large language model interprets the data and recommends intervention methods based on the most prevalent project risk at any given time.

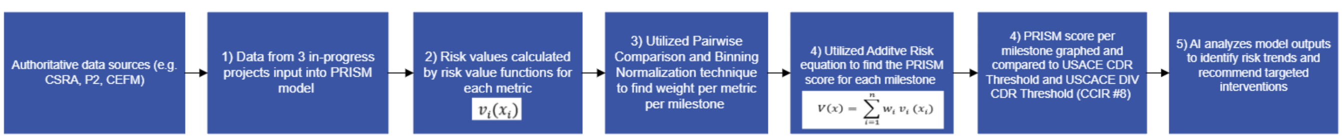

Figure 3: Methodology

To calculate the PRISM score, the analysis assigns a weight to each risk measure through a pairwise comparison, evaluating its importance relative to project health. The team conducted a pairwise comparison for each life cycle phase. The model generates risk functions, also known as value functions, for each risk measure to show how their contribution to overall risk changes across the range of values for that measure. For example, SPI has an exponentially decreasing value function on a range of 0.2 to 1.8. An SPI of 0.2 contributes the maximum possible risk for that metric, an SPI of 1.0 contributes only 20% of the risk as an SPI of 0.2, and an SPI of 1.8 contributes zero risk. To calculate the PRISM score, the team converted the raw data into risk values based on the risk functions and computed a weighted average across all seven-risk metrics.

Figure 4 shows the output of the pairwise comparison of all risk metrics by phase. The pairwise comparison applies fundamental project management principles to determine which metrics contribute the most risk relative to the others. The result is a set of swing weights that indicate the impact of each metric on the overall PRISM score for that phase. When raw data enters the PRISM model, the system assigns a risk score to each data metric based on its associated risk function. The model then multiplies that risk score by the metric's swing weight for the given phase. The sum of all weighted risk scores produces the PRISM score for that milestone. Lower PRISM scores indicate less risk while higher scores signal increased risk.

Figure 5 presents the final numeric output of the PRISM model, displaying PRISM scores by milestone for a specific project, along with the defined risk threshold for each phase.

Swing Weights By Phase								
Phase	CPI	SPI	% Contingency	# VE Adjustments	# REQ Changes	GC Score	Risk-Liklihood	total
Initiation	0.071	0.071	0.128	0.231	0.231	0.142	0.125	1
Planning & Design	0.218	0.218	0.091	0.134	0.134	0.121	0.083	1
Execution & Control	0.274	0.298	0.088	0.116	0.116	0.063	0.045	1

Figure 4: Weighting Process for Metrics

PRISM			
Phase	Milestone	Progress	Score
Project Initiation Phase	0-30% Design Development	0-30%	13.55
	60-90% Design Development	60-90%	13.40
	90-Final% Design Development	90-100%	36.56
Project Planning & Design Phase	1391 First Draft	0-50%	25.74
	1391 Final Draft	50-100%	28.33
Project Execution & Control Phase	0-30% Design Constructed	0-30%	21.26
	30-60% Design Constructed	30-60%	23.26
	60-90% Design Constructed	60-90%	55.12
	90-MILCON Project Fully Constructed	90-100%	0.00

Figure 5: PRISM Score Matrix

3. Findings

3.1 Model Validation

Figure 6 provides a visual representation of the PRISM score over time for three separate projects. For visualization purposes, the Veterans Integrated Services Network (VISN) commander sets the same risk threshold for all three projects. VISNs are regional systems within the VA that manage healthcare operations across multiple VA medical centers. The stair-step pattern in the graph illustrates how risk tolerance increases as a project progresses. The threshold represented in Figure 6 is a proxy threshold set for demonstration purposes. For future client use of the PRISM model, senior leaders will identify and decide upon this threshold through analysis of historical project data.

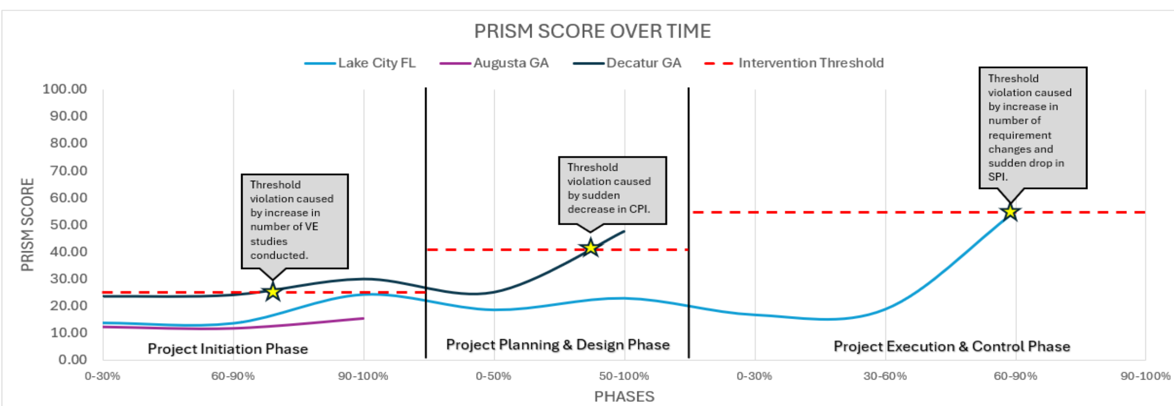

Figure 6: PRISM Score Over Time for Three Projects

Higher risk at the beginning of a project often leads to greater overall project risk. Viewing PRISM scores in this manner allows project managers to assess whether project risk is increasing or decreasing over time, how quickly it is changing, and—most importantly—when it exceeds an acceptable threshold. When a project's PRISM score crosses the red dotted line representing the risk threshold, it triggers an intervention. Analyzing the AARs of the three projects shown in Figure 6 offers further insights into the PRISM methodology. The Augusta project currently has minimal risk, but the team expects it to experience a significant spike soon due to imminent requirement changes. A hurricane severely affected the Decatur, GA project, causing ongoing supply chain disruptions and significant schedule delays. Of the three projects, this one is in the worst condition based on both AARs and PRISM scores, which accurately reflect its heightened risk as it remains near the threshold from the beginning of the project. The Lake City, FL project is experiencing contract disputes and steelworker strikes, leading to supply shortages late in project execution. The PRISM score reflects this by spiking during the Execution & Control Phase. Overall, comparing PRISM scores with project AARs has confirmed the model's effectiveness in predicting project risk.

3.2 Sensitivity Analysis

A tornado diagram shows how individual metrics affect the overall PRISM score by visually representing sensitivity of each project's seven metrics in changes to the various input variables. Historical data from ten completed projects determined the high and low values of weighted risk. To measure each metric's effect, the calculation of a PRISM score involved adjusting its high and low values while keeping the average weighted risk of the other six metrics constant. A tornado diagram highlights which metrics cause the greatest variation in PRISM scores, and thus VA leadership could use this to prioritize efforts on focused mitigation strategies for the most critical risks. Figure 7 shows that CPI and SPI, the two earned value management metrics, exert the most influence on the PRISM score during the Project Execution & Control Phase. Further analysis of earlier project phases would reveal greater variation from other metrics, as CPI and SPI gain more influence in later stages of the project. The takeaway is that the attributes with the greatest swing weights have the strongest influence on the PRISM score.

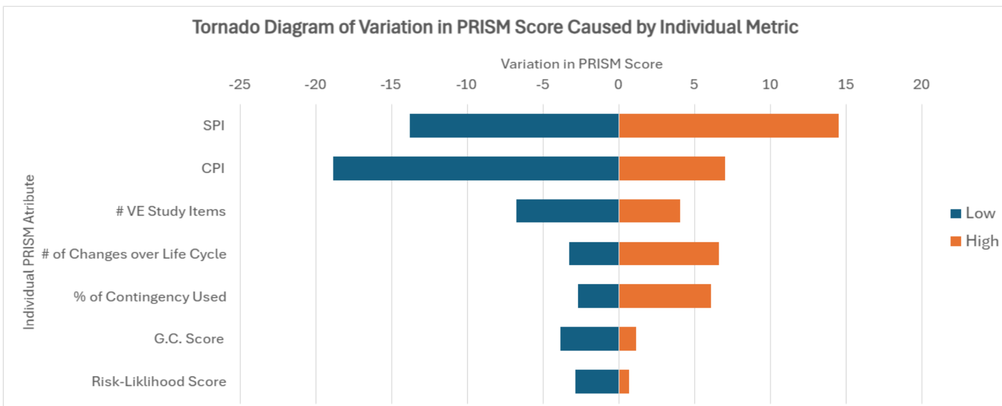

Figure 7: Individual PRISM Metric Sensitivity Analysis (Execution & Control Phase)

4. AI Validation and Testing

Project teams can expand the utility of the PRISM model by integrating AI to support early risk detection and mitigation, building on its data-driven method for evaluating project health. The challenge lies in deciding whether AI can reliably interpret project performance metrics and offer sound recommendations across key areas such as cost, schedule, and scope. To confirm AI's potential in project management decision-making, the team assessed ChatGPT 4o, OpenAI's most advanced model, against a practice PMP test. This test aligns with earlier work from Koban et al. (2025) and Yeo et al. (2023) in assessing AI's performance on professional certification examinations. This assessment answered two research questions: 1) How accurate and complete are ChatGPT 4o's responses to practice PMP test questions?, and 2) Can CFM confidently use AI for aid in making high-level project management decisions?

The practice test, sourced from a reputable PMP test prep company, consists of two hundred questions categorized into the following categories: human resources, cost, schedule, and scope. The team input each question into ChatGPT 4o, prompting the model to generate an answer with an explanation. Once complete, the team evaluated the responses against the answer key and categorized them based on Koban et al.'s assessment criteria: 1) Comprehensive: Correct answer with a correct explanation, 2) Correct but inadequate: Correct answer but weak or incorrect explanation, 3) Mixed information: Incorrect answer with partially correct or misleading reasoning, and 4) Completely incorrect: Incorrect answer and explanation.

Results showed high accuracy, with scheduling questions scoring the lowest at 91%. Across all two hundred questions, 80% were both correct and well-explained, while the remaining responses held minor inaccuracies or inadequate justifications. For VA capital investments, executive decision-makers should prioritize cost, schedule, and scope when considering AI's role in project management. While AI provides valuable support, these findings emphasize the need for human validation to ensure accuracy. As AI technology continues to evolve, its application in risk assessment and intervention strategies will improve, creating new opportunities to enhance project oversight in the VA.

Category	Total Questions	Correct Answers	Accuracy
HR	55	53	96%
Cost	33	32	97%
Schedule	43	39	91%
Scope	69	66	96%

Explanation Quality	Count	
1. Comprehensive	160	Correct
2. Correct but inadequate	25	
3. Mixed	10	Incorrect
4. Completely incorrect	5	

Figure 8: AI Performance Metrics

6. Conclusions and Future Work

This study introduced the PRISM framework, a novel approach to project risk assessment and mitigation for VA medical facilities. By integrating systems engineering methodologies with milestone-based risk metrics, the PRISM model provides project managers with a real-time, data-driven tool for monitoring project health and predicting risk trends. Our findings demonstrate that PRISM offers a structured and transparent way to quantify project risk, allowing for early intervention and improved decision-making. The framework aligns with industry best practices and enhances stakeholder understanding of project health, making it a significant step forward in modernizing risk management for VA construction projects.

However, our research relied on several limitations that must be addressed in future work. The model's validation was based on data from a limited sample of three VA projects, which constrains the generalizability of our findings. Another key limitation is the model's inability to process unstructured AARs effectively, which limits its ability to capture risk factors beyond structured cost and schedule metrics. Furthermore, while AI was tested on PMP exam questions to validate its understanding of project management principles, its practical application in real-world risk assessment still requires further refinement and testing.

To build upon this foundation, future work should focus on expanding the dataset to include a broader range of VA projects, ensuring the model's robustness and scalability. A major next step involves developing a data strategy to transform unstructured AARs into structured inputs, enabling AI-driven analysis of historical project risks. Finally, integrating an AI-enhanced PRISM dashboard will provide project managers with an interactive tool for monitoring project health, evaluating risk, and receiving tailored intervention recommendations. By addressing these areas, PRISM has the potential to become a transformative solution for improving risk management and decision-making in VA construction and facilities management.

7. References

Batt, G., Kusano, M., McMickens, T., Rubin, J., & Schreiner, J. (2024). *Priority Risk Index: A novel approach to determining project health for the United States Corps of Engineers*. Department of Systems Engineering, United States Military Academy.

Cost Estimating Requirements for Veterans Affairs Facilities; Office of Construction and Facilities. (2022, August 1). *VA.gov Veterans Affairs*. Retrieved October 2, 2024, from https://www.cfm.va.gov/cost/index.asp

Dexnova Consulting Limited. (2013). *PMP practice questions and answers.* Dexnova Consulting Limited.

Fahad, U. PMP. (2024, March 15). What is a Cost Baseline in Project Management? | PM Study Circle. PM Study Circle. https://pmstudycircle.com/cost-baseline/

Koban, D. D., et al. (2025). ChatGPT as a study tool: Exploring its potential to help industrial and systems engineering students prepare for the FE exam [Unpublished manuscript]. United States Military Academy, West Point.

OpenAI. (2024). ChatGPT (April 5 version) [Large language model]. https://chat.openai.com

Parnell, G. S., Driscoll, P. J., & Henderson, D. L. (2011). *Decision making in systems engineering and management* (2nd ed.). John Wiley & Sons.

U.S Army Corps of Engineers Cost Engineering Center of Expertise. (2024, March). *USACE Recommended Best Practices Cost and Schedule Risk Analysis (CSRA).* USACE Recommended Best Practices, Cost & Schedule Risk Analysis (CRSA).

U.S. Army Corps of Engineers. (2018). USACE Business Process (ER 5-1-11). Department of the Army ER 5-1-11, USACE Business Process.

Y.H. Yeo et el., "Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma," Clin. Mol. Hepatol., vol. 29, no. 3, pp. 721-732, Jul. 2023, doi: 10.3350/cmh.2023.0089.

ISBN: 97819384969-9-8