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

A Systems Decision Process Approach to Lockheed Martin's Human in the Loop Study Design

Matteo Finn, Austin Kiyaani-McClary, Erich Marschall, Alexander Vollmond, Patrick Mugg, and Chadwick Clark

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

Corresponding author's Email: erich.a.marschall@gmail.com

Author Note: We extend our deepest appreciation to Lockheed Martin Senior Fellow Mike Skaff, Associate Fellow Keith Higginbotham, and Dave Wozniak for their exceptional collaboration throughout this process. Our heartfelt thanks go to LTC Mugg and Dr. Chadwick Clark for providing this remarkable opportunity to expand our knowledge, develop as students, and enhance our research capabilities. The views expressed in this paper represent those of the authors and do not reflect the positions of the United States Army, the United States Military Academy, or the Department of Systems Engineering.

Abstract: This research utilizes the Systems Decision Process (SDP) to improve Lockheed Martin's Human-In-The-Loop (HITL) studies for usability evaluation of the current Pilot-Vehicle Interface (PVI) in the F-35 Lightning II. This study focuses on the first three phases of the SDP, from problem definition to decision-making to show that a consistent study design is optimal for usability evaluation. The authors generated a quantitative value model to assess the usability of the candidate solutions: slingshot, single tap, double tap, quick action, button, and long press actions on the PVI. To quantify naturalness, workload, feedback, and efficiency, an additive value function generated a value matrix. Additionally, the value matrix incorporates notional biometric data to create weighted value scores. Using this data, the generated model compares alternative candidate solutions of each value measure. Our findings serve as an alternative to the Bedford Workload Rating Scale (BWRS), Likert scale, and other methods of usability analysis.

Keywords: Human in the Loop Study, Usability, Systems Decision Process, Pilot-Vehicle Interface, Pilot Sensing Technology, Collaborative Combat Assistants

1. Introduction

Defense industry research efforts work toward developing Artificial Intelligence programs that integrate machine learning and pilot calibration to reduce hazards linked to human error. In 2020, the National Commission for Military Aviation Safety reported that military aviation accidents increased from 2013 to 2018, with 43% attributed to human error (NCMAS, 2020). Factors, such as the design and usability of cockpit interfaces can significantly increase pilots' cognitive load, confusion, and operational difficulty. Poor design and usability can overburden the pilots and increase the likelihood of human error. Next generation aircraft designs seek to embed teams of unmanned Collaborative Combat Aircraft (CCAs) with pilots who can control these assistants to combat enemy aircraft (Oxford Analytica, 2023).

The importance of monitoring workload and usability grows with the increasing complexity of combat in the air domain. Lockheed Martin has found that biometric data collected with Pilot Sensing Technology (PST) provides valuable insights in potential error-prone scenarios (Carlsen et al., 2024). Their HITL studies regularly use PST to track biometric data. These studies assisted in the development of an individualized machine learning program that compares a pilot's biometric baseline to any deviations to classify if the pilot is experiencing high or low Mental Workload (MWL) (Carlsen et al., 2024). Pilots complete assigned tasks by interacting with pilot-vehicle interfaces (PVI) that act as the cognitive, sensory, and psychomotor link between the pilot and onboard systems, including CCAs (Fowler & Rogers, 1991). Poorly designed PVI can overwhelm pilots with excessive information, inaccessible controls, or unclear layouts, significantly increasing cognitive workload and stress. Multitasking in such environments amplifies these challenges, making the relationship between pilots and cockpit technology a critical factor for safety. Consequently, Lockheed Martin now seeks to evaluate the usability of different PVI design features by examining pilot input and MWL derived from biometric data trends collected by PST. To address this objective, we used the SDP to evaluate various methods for designing a PVI.

Lockheed Martin's most recent HITL studies put pilots into air-to-air combat scenarios where they controlled CCAs via in-cockpit PVI to destroy enemy combatants. These studies used a F-35 cockpit with a tablet that provide pilots a menu and screen to control their CCAs. Lockheed Martin provided post-study surveys, and conclusions from five HITLs that shared this general design. Aside from this design, we found that the HITLs varied greatly. With study participation ranging from 1 to 6 participants, studies were used to evaluate PVI design and PST implementation for data collection optimization. On several occasions, test subjects were involved in the design or execution of other HITLs before participating in one themselves. With the heterogeneous nature of the biometric data and survey responses, we decided to conduct a case-study on how Lockheed Martin conducts their HITLs. Ultimately, our goal is to meet their intent by generating an alternative method to their HITL design.

2. Problem Definition

The SDP, shown in Figure 1, guides activity, planning and thinking for complicated projects (Driscoll et al., 2022). We used this process to conduct our case study and offer Lockheed Martin an alternative for HITL design to conduct effective usability evaluation for next generation aircraft.

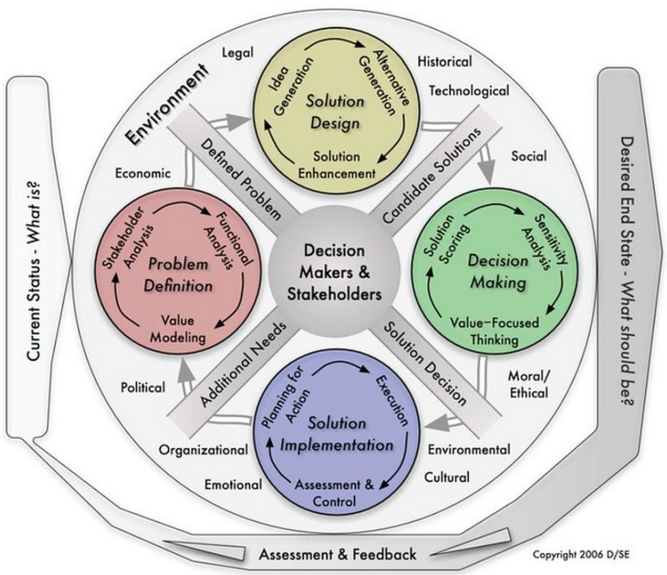

Figure 1: Systems Decision Process

Our initial problem statement for this capstone research project was that the Pilot-Vehicle Interface of Lockheed Martin's F-35 must enable seamless communication between pilots and aircraft systems in high-stress environments, minimizing workload and maximizing operational efficiency. The premise of our analysis assumes that assessing the impact of PVI on pilot workload can improve interface design, understanding of pilot performance, and decision making under pressure.

We conducted a Findings, Conclusion, and Recommendations (FCR) matrix to summarize our initial research and stakeholder analysis. The FCR meets this objective by listing significant facts discovered during stakeholder analyses and research, the team's interpretation of fact relevancy, and follow-on actions that the team recommends or intends to follow in subsequent SDP phase activities (Driscoll et al., 2022). Our FCR, relied on insights gained from reviewing relevant literature, interviews with Lockheed Martin's HITL Lab research scientists, and our visit to the F-35 production facility. While our initial focus was on the e-Pilot program as a whole, we refined our scope to specifically evaluate the design of the HITL studies. After conducting site visits and lit reviews, we iteratively developed the following problem statement: Lockheed Martin shall optimize human-in-the-loop study designs to evaluate the usability of Pilot-Vehicle Interface features in the F-35, ensuring natural interaction, task efficiency, workload reduction, and effective feedback under simulated flight conditions using representative F-35 cockpit environments and pilot test subjects.

ISBN: 97819384969-9-8

As part of the problem definition phase of the SDP, the team leveraged a value-modeling approach that incorporates both qualitative and quantitative data to inform the design process (Driscoll et al., 2022). The qualitative value model develops and refines a set of objectives for a system in a hierarchy-style depiction as shown in Figure 2.The primary level of the hierarchy is the fundamental objective, or purpose of the project. Our fundamental objective is to optimize the HITL study design to evaluate usability. The second tier of the hierarchy includes the functions that will fulfill the fundamental objective. The last tier of the hierarchy includes the value measures as quantifiable metrics to assess progress made for each objective. For each value measure we determined a metric to classify performance where More is Better (MIB) or Less is Better (LIB). We want PVI designs to help complete tasks, minimize stress, and inform pilots. Each function has a set of objectives to either maximize or minimize to meet the intent of the function. Naturalness refers to how intuitive and physically natural a task feels to the pilot when using a particular PVI. It is measured through pilot survey results on a 1-7 likert scale and gaze entropy. Efficiency is defined by maximizing the effort required to achieve an objective, assessed by the number of steps needed to access a specific CCA menu and feedback from pilot surveys. Workload is measured through pilot survey responses, HRV recorded in beats per minute, and PCPS measured as a percentage. Finally, feedback is also measured solely through pilot survey results on a 1-7 likert scale where 7 is the best.

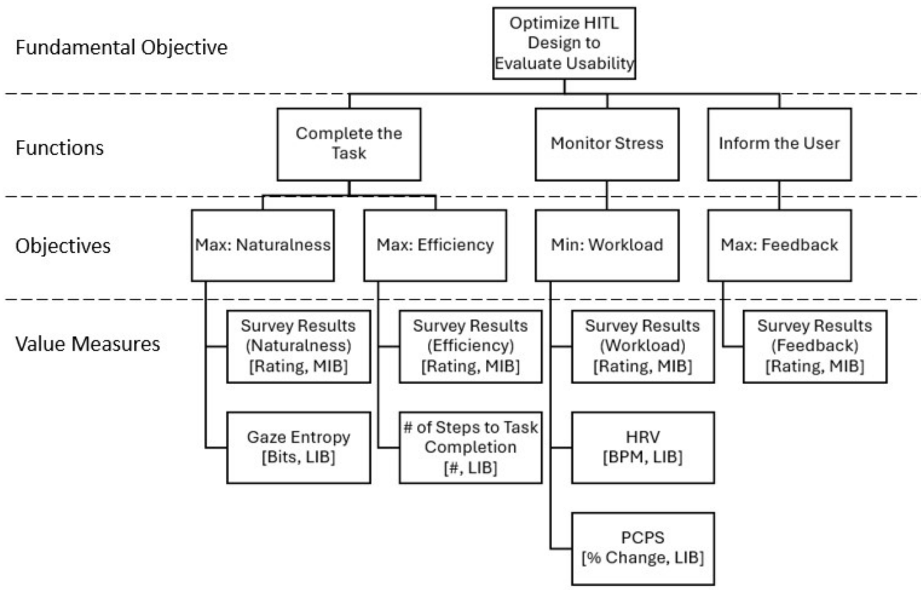

Figure 2: Qualitative Value Model for HITL Design Optimization

A value matrix converts qualitative criteria from the qualitative value model into measurable values, ultimately developing candidate solution evaluation criteria. First, decision makers assigned numerical values to various value measures that determine how well a candidate solution reaches an objective. Decision makers develop these measures based on their alignment with the objective and their scale of measurement. (Driscoll et al., 2022) From our qualitative value model, our value measures include gaze entropy, number of steps to task completion, Heart Rate Variability (HRV), Percent Change Pupil Size (PCPS), and pilot feedback survey results based on naturalness, efficiency, workload, and feedback. During the development of these value measures, independent and dependent variables are assigned. The independent variables represent the input being evaluated, while the dependent variables indicate the value output or the importance of a particular input to the stakeholder. Many of the numerical values assigned for each value measure are notional and subjective. The only value measures that are not notional are the feedback survey results since our team received qualitative survey data from Lockheed Martin. To determine the overall value for each value measure, we identified the swing weights and determined the global weights for each value measure via stakeholder input. Statistical analysis typically determines variable importance. Given the lack of experimental data, we opted to assess variable importance with survey feedback from Lockheed Martin analysts about their subjective views of each measure. We interviewed three researchers who helped design HITL studies and assist in the creation of MWL classifiers using

ISBN: 97819384969-9-8

in-cockpit PST (Carlsen et al., 2024).

We collected the majority of information on swing weights from a Human Factors Design Engineer who conducted initial data analysis and collection on all five rounds of HITL studies that Lockheed Martin has conducted to date. We asked for how swing weights should be prioritized and what numerical values should represent importance. We also asked for appropriate measurements and ranges for the four objectives. These personal assessments of swing weights were approved by the former F-16 pilot and current Chief Engineer of PVI for the F-35 Program and a Human Performance Engineer with over 10 years of experience. Out of all the value measures, Lockheed Martin emphasized that feedback holds the greatest importance, therefore, it has a swing weight of 100. All other value measures are then assigned a swing weight value on a scale from 0 to 100 in comparison to the most important value measure. For example, number of steps to task completion having a rating of 60 means that it is 60% as important as feedback. The higher the value assigned to the value measure, the greater the importance to Lockheed Martin. To calculate the global weights, the individual swing weight's are divided by the sum of the total swing weights. Once the global weights are found as shown in Figure 3, we utilized the additive value model from Equation (1) in the next phase of the SDP to determine the overall value for each PVI design alternative.

Value Measure	Swing Weights (f_i)	Global Weights (w_i)
Heart Rate Variation (HRV)	30	0.100
Eye Movement (% Change)	30	0.100
Steps to Task Completion (#)	60	0.200
Feedback (Survey Results)	100	0.333
Gaze Entropy (Bits)	20	0.067
Naturalness (Survey Results)	10	0.067
Efficiency (Survey Results)	30	0.100
Work Load (Survey Results)	20	0.033

Figure 3: Additive Value Model Weights

3. Solution Design

During the solution design phase of the SDP, our team explored the PVI design alternatives that Lockheed Martin generated. Lockhee Martin's goal is to determine which PVI designs to implement in F-35 cockpits, focusing primarily on PVI usage for controlling CCAs. Our data analysis incorporated survey feedback from pilots, emphasizing their preferences. The qualitative analysis came from pilot comments on specific aspects they liked about the PVI. Lockheed Martin's design alternatives include single tap, double tap, button, quick action, slingshot, and long press.

Alternatives	Brief Description
Single Tap	When selected with a single tap, the CCA expands to 150% in size with an orange outline on the PVI. The assigned target or location is then selected by single tapping on the PVI.
Double Tap	When selected with a double tap, the CCA expands to 150% in size with an orange outline on the PVI. The assigned target or location is then selected by double tapping the PVI.
Quick Action	When selected with a double tap or single tap, the CCA expands to 150% in size with an orange outline on the PVI. The assigned target or location is then selected by double tapping or single tapping the PVI.
Long Press	When selected with a single tap, the CCA expands to 150% in size with an orange outline on the PVI. The assigned target or location is then selected and confirmed by long pressing the PVI.
Button	Buttons, labeled as option selection buttons, allow pilots to select CCAs and confirm location changes.
Slingshot	The slingshot method involves dragging a selected CCA in the opposite direction of the intended heading, with orange arrows indicating the new trajectory. When the pilot releases their finger from the PVI, this commands the CCA to change heading.

Figure 4: Alternative Descriptions

Lockheed Martin compared PVI features in pairs using survey feedback and a Likert scale (no preference = 0, somewhat prefer = 1, prefer = 2, strongly prefer = 3) to determine pilot preferences. While this design parameter efficiently eliminates less favored alternatives, our team's model took a more holistic approach by evaluating value measures that contribute to each PVI alternative. Our approach accounts for additional factors that influence implementation decisions such as considering gaze entropy, eye movement, and heart rate compared to utilizing just survey results which is what Lockheed Martin utilized for determining their implementation. To determine the overall value for each design alternative, we utilize the additive value model referenced in Equation (1), where w_i is the global weights, $v_i(x_{j,i})$ is the value score of the j^{th} alternative from the i^{th} value measure. The additive value model uses the global weights and value scores from Figure 5 for each PVI design alternative. The model then computes a total value score for each design alternative as seen in Figure 6, which is a stacked bar chart.

$$V(x_j) = \sum_{i=1}^{n} [w_i \times v_i(x_{j,i})] \tag{1}$$

	Value Measure ($v_i(x_{j,i})$)							
Candidate Solution (j^{th} alternative)	Heart Rate	Eye Movement	# Steps to Task Completion	Feedback	Gaze Entropy	Workload Survey	Effeciency Survey	Naturalness Survey
Single Tap	60	65.0	100	30	80	100	95	50
Button	62.5	33.8	100	30	3	80	20	80
Quick Action	93	46.0	60	15	75	50	80	50
Slingshot	50	76.7	80	100	96	50	95	100
Double Tap	97.25	97.6	20	30	60	50	50	80
Long Press	32	61.7	80	30	25	80	80	95

Figure 5: Value Scores for HITL Design Alternatives

4. Decision Making

In the Decision Making Phase of the SDP, the primary objective is to optimize the HITL study design for usability evaluation. Utilizing stakeholder feedback and subjective assessments, the qualitative value model generates a weighted value stacked bar chart for each alternative. As shown in Figure 6, each alternative's weighted values are represented by different colors corresponding to specific value measures. The ideal alternative is a snapshot of a potential alternative with the maximum value for each value measure. The team then compared the different PVI alternatives against the ideal alternative to determine which solution is best with an understanding of the factors that play the largest role in their score. The results indicate that the Slingshot alternative is the most favorable for Lockheed Martin. The chart highlights the highest and lowest contributing factors for each candidate solution, enabling improvement and trade-off analysis for any alternative. Lockheed Martin can use these insights to refine their HITL study design, increasing the value of preferred alternatives. Additionally, they can develop a more optimized and consistent test study design for future HITL studies, improving multiple defined value measures.

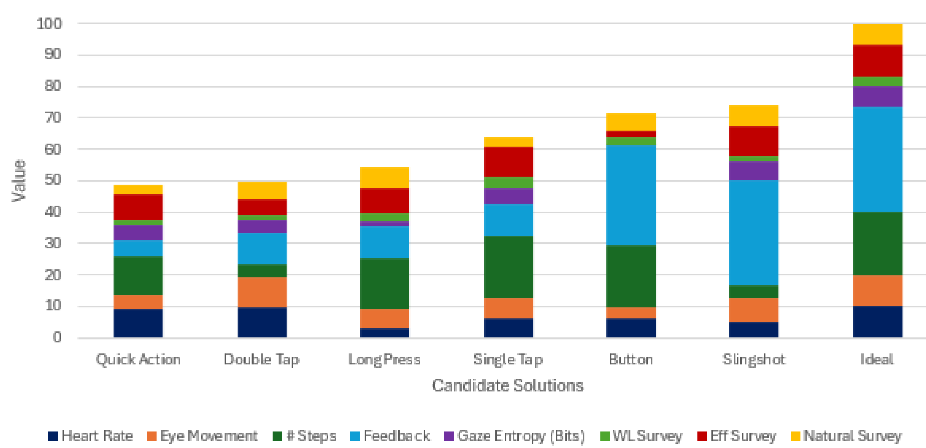

Figure 6: Value Scores of PVI Alternatives

5. Conclusion

Future work could compare these conventional methods with value matrices built using the additive value function and stakeholder feedback to determine which methods provide the most desirable feedback for stakeholder requirements. Understanding how these different approaches influence decision-making will help refine methodologies and ensure that future studies are designed to capture the most relevant insights.

Our research is designed to improve Lockheed Martin's PVI usability analysis, but our approach has potential for widespread applicability. The additive value model can quantitatively score any qualitative attribute. Usability studies for other forms of burgeoning technology like AI assistance and autonomous vehicles can apply the additive value model to compare and contrast various designs. The SDP which includes the additive value model, sensitivity analysis and other rigorous assessments is also applicable to research topics in the civilian and defense sectors. The SDP framework and additive value model we used is suitable for any research that seeks to quantify measures of qualitative attributes for multiple alternatives.

Future research should also consider how this approach fits within the Department of Defense acquisition process since acquisition constraints and technical feasibility will equally influence implementation. Rewriting several lines of code may seem intuitive, but solution implementation is an intricate process. Moving from research to operational capability involves multiple stages: identifying the requirements, engineering development, testing, and post-deployment. Each stage has its own challenges including strict budget cycles, long procurement timelines, and the need to integrate with existing defense programs. These intricacies make adopting new PVI solutions a lengthy and costly process. Future research will need to explore the navigation of regulatory requirements, securing of funds, and project alignment with key acquisition milestones that are prerequisites to implementation in future aircraft systems. (Small Business Administration, 2025)

The cost of implementing new aviation technologies within the military is also a significant factor to consider for solution implementation analysis. In the Fiscal Year 2025 budget, the Department of Defense allocated \$61.2 billion which is approximately 20% of the total investment budget—to aviation and related systems (Office of the Under Secretary of Defense (Comptroller), 2024). While this figure highlights the strategic importance of aviation, it also underscores the financial constraints and competing priorities within the broader defense budget. Programs such as the F-35 Lightning II alone account for \$12.4 billion, demonstrating the high costs associated with modernizing military aviation (Office of the Under Secretary of Defense (Comptroller), 2024). Given these expenditures, any new PVI solution not only needs to demonstrate operational benefits but also justify its cost-effectiveness within the acquisition landscape where funding is heavily directed toward shipbuilding, missile defense, and space-based systems. Successfully integrating new usability improvements will require careful navigation of budget cycles, prioritization within program funding, and a clear demonstration of return on investment in terms of both mission effectiveness and life cycle costs.

References

Carlsen, V., Manzi, R. E., Dellinger, S., Craig, T., & Koban, D. (2024). *Predicting pilot workloads using physiological measures.* Retrieved from `https://www.iser.sisengr.org/index.php/iser/article/view/202` (White Paper)

Driscoll, P. J., Parnell, G. S., & Henderson, D. L. (2022). *Decision making in systems engineering and management* (3rd ed.). Wiley.

Fowler, T. D., & Rogers, S. P. (1991). *A clips-based tool for aircraft pilot-vehicle interface design.* Retrieved from `https://ntrs.nasa.gov/api/citations/19920007387/downloads/19920007387.pdf`

NCMAS. (2020). *Military aviation losses fy2013 - fy2020: 224 lives | \$11.6 billion | 186 aircraft.* Retrieved from `https://www.globalsecurity.org/military/library/report/2020/ncmas_final_report_20201201.pdf` (Accessed: 2023-03-06)

Office of the Under Secretary of Defense (Comptroller). (2024). *Fiscal year 2025 budget request: Program acquisition cost by weapon system* (Budget Report). U.S. Department of Defense. Retrieved from `https://comptroller.defense.gov/Portals/45/Documents/defbudget/FY2025/FY2025_Weapons.pdf` (RefID: 9-E6AAC8A)

Oxford Analytica. (2023). *Crewed and uncrewed combat aircraft will collaborate.* `https://doi.org/10.1108/OXAN-DB282804`. (Expert Briefings)

Small Business Administration. (2025). *Department of defense (dod) acquisition basics.* Retrieved from `https://www.sbir.gov/tutorials` (Accessed: March 2025)