*Human Factors in Design, Engineering, and Computing,* Vol. 159, 2024, 1255–1265
https://doi.org/10.54941/ahfe1005693

# Analysis of Pre-Flight and Monitoring Tasks Using Cognitive Performance Modeling

**Chihab Nadri[1], Yunmei Liu[2], Maryam Zahabi[1], David Kaber[3], Jaime Ruiz[3], Michael Middleton[4], and Ryan McKendrick[4]**

[1]Texas A&M University, TX, USA
[2]University of Louisville, KY, USA
[3]University of Florida, FL, USA
[4]Northrop Grumman Corporation, VA, USA

## ABSTRACT

Pilot cognitive workload and errors significantly contribute to aviation incidents. Evaluations of piloting tasks in both simulators and real-world settings, along with computational models of cognitive task performance can help to identify cognitively challenging tasks early in the system design process and enhance user interface designs. This study applied cognitive performance modeling (CPM) to assess pilot task demands in pre-flight and monitoring using a UH-60V Black Hawk helicopter flight simulator. The objective was to propose potential flight checklist and subtask interface redesigns to reduce pilot working memory load and improve operational effectiveness. Initial analysis involved reviewing pilot instructions and logs for pre-flight checks, monitoring activities, and emergency responses. Actions, such as button presses, task errors and the duration between tasks were recorded. A Hierarchical Task Analysis (HTA) was applied to identify sub-task interdependencies. CPMs were developed using Cogulator and a variation on the GOMS language detailing cognitive, perceptual, and motor processes. Models focused on task sequences and cognitive process durations, revealing task time estimates, working memory load, and cognitive workload. Demanding subtasks were identified based on longer durations and/or higher workloads. Cogulator model outcomes for workload assessment were compared with pilot opinions on task difficulty for model validation. Recommendations for cockpit interface enhancements were formulated with the aim of streamlining sub-task operation sequences, reducing cognitive load, and improving pre-flight and monitoring efficiency. Key suggestions included redesigning checklists, providing auto-text completion options for data entry tasks, and implementing temporary shutdowns of displays (irrelevant to the primary flight task) under emergency conditions. The study methodology was validated through expert interviews and findings inform the design of current and future piloting procedures, potentially contributing to improved aviation safety and efficiency.

**Keywords:** Cognitive performance modeling, Cognitive workload, Aviation human factors, Occupational safety

## INTRODUCTION

Cognitive workload is a product of human information processing limitations and can lead to performance problems, particularly in high-stakes environments such as aviation. For pilots, cognitive workload can create flight safety risks as it can negatively affect situation awareness and decision making (Amalberti & Wioland, 2020; Wiegmann & Shappell, 2001). Pilots must constantly respond to and remember vast amounts of information from multiple sources. Cognitive overload and loss of situation awareness has been associated with performance degradations, including piloting (Onnasch et al., 2014; Jones & Endsley, 1996).

The UH-60 Black Hawk helicopter is a military aircraft with multiple displays and interfaces for pilots to monitor in flight (Figure 1). The complexity of the aircraft cockpit imposes high information processing demands for pilots and necessitates advanced training and cognitive support systems to reduce the likelihood of errors. As such, the Black Hawk has been involved in a number of incidents attributed to pilot cognitive overload. Since 2020, there have been 31 crash incidents (Aviation Safety Network, 2024). As an example, on April 6th, 2023, Japanese pilots encountered an engine emergency that led to a fatal crash, killing all ten people on board. Cockpit voice recordings revealed the crew struggled in managing the loss of engine power and corresponding warnings, ultimately becoming overwhelmed with the situation.

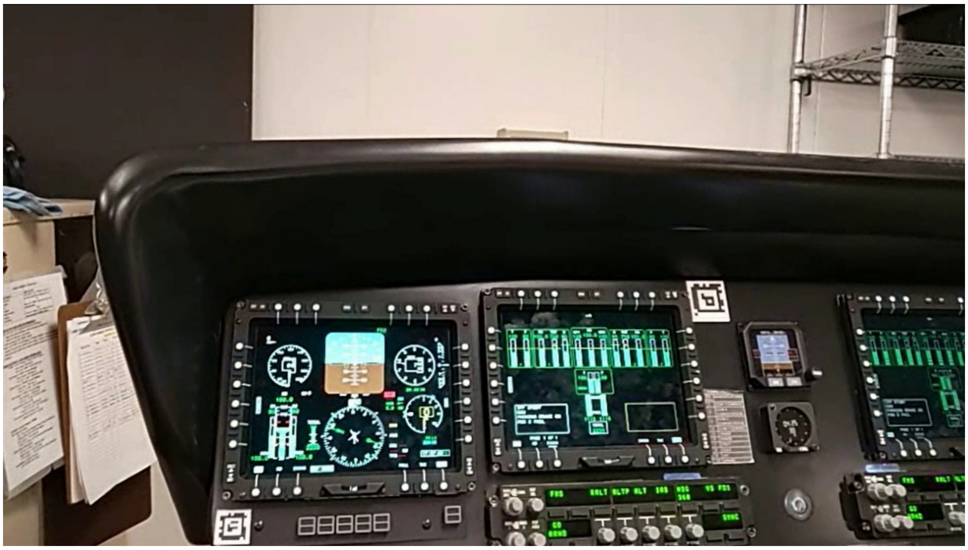

**Figure 1:** View of UH-60V Black Hawk helicopter simulator.

As the above incident (and others) underscore the potentially dangerous impact of pilot cognitive overload, researchers have conducted human factors evaluations of piloting tasks using simulators and field testing to understand and attempt to mitigate flight risks (Fujizawa et al., 2014). These evaluations (e.g., Goodwin, 2017; Havir et al., 2006; Thornton et al., 1992) have identified training effectiveness, crew coordination, and crew station suitability to task as contributing factors. One promising approach to human factors evaluations is cognitive performance modeling (CPM). CPM enables researchers to identify cognitively challenging tasks early in the system design process (Park & Zahabi, 2024). By doing so, these models also facilitate user interface design enhancements and pilot training procedures towards reducing cognitive workload (Wickens, 2002). A critical component of effective pilot training is the design of flight checklists and standard operating procedures (SOPs). Checklists are essential tools that help pilots manage tasks systematically and consistently, ensuring that no critical steps are overlooked (Gawande, 2010). As pilots frequently refer to checklist items in regular flight operations, the lists represent a useful reference for CPMs of flight tasks.

In this study, we focused on modeling Black Hawk pre-flight and monitoring task cognitive demands. Our objective was to identify high workload subtasks in pre-flight and monitoring and generate flight checklist and subtask redesigns to reduce pilot working memory load and promote operational effectiveness. Initial analysis involved a comprehensive review of pilot instructions and using hierarchical task analysis (HTA) to formally represent pilot plans, task sequences, decision requirements and subtask dependencies. The HTA was then used along piloting logs from an experiment with a UH-60V simulator to develop CPMs of pre-flight and monitoring task operations.

## METHODS

### Hierarchical Task Analysis

For the HTA, each phase of flight (pre-flight and monitoring) was systematically divided into functional tasks (see Figure 2, Tables 1 and 2), contained subtasks and individual operations (see Figure 2 as example).

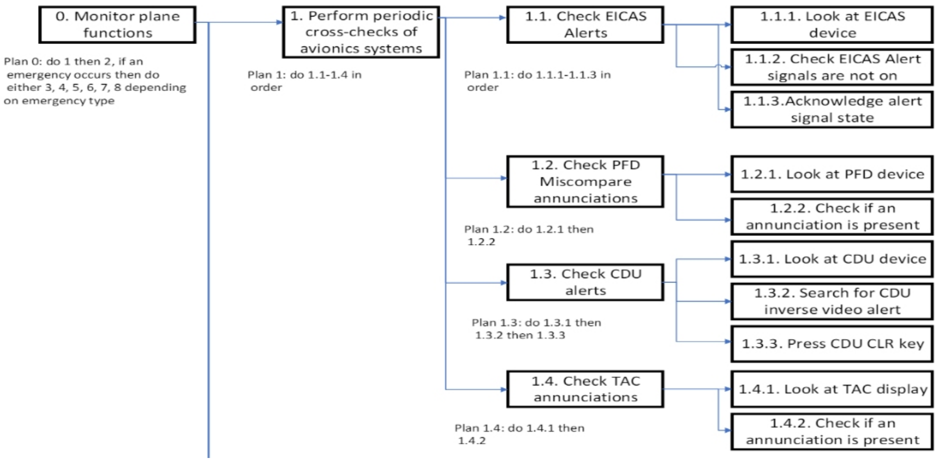


**Figure 2:** HTA breakdown of monitoring task 1.

**Table 1.** Pre-flight task breakdown.

| Pre-Flight Task # | Description |
|---|---|
| 1 | Load mission data |
| 2 | Confirm NAV DB / TAC DB |
| 3 | Check Avionics |
| 4 | FMS Initialization |
| 5 | Check EICAS Alerts |
| 6 | Check CDU Alerts |
| 7 | Set Up MFD/FD/FMS for Mission |
| 8 | Verify Barometric Altimeter |
| 9 | Notify Crew of GO/NGO ready for takeoff |

**Table 2.** Monitoring task breakdown.

| Monitoring Task # | Description |
|---|---|
| 1 | Perform periodic cross-check of avionics systems and instruments. |
| 2 | Monitor Route / Flight Plan parameters |
| Emer-1 | Critical data miscompare via Miscompare Annunciation / MSG annunciation |
| Emer-2 | Respond to EICAS Alerts |
| Emer-3 | Respond to CDU Msg / Alerts |
| Emer-4 | Pop-up Weather Event |
| Emer-5 | Engine / Fuel Emergencies |

For example, a monitoring systems check is broken-down into subtasks, including checking avionics systems. Each of these subtasks is then further decomposed into specific actions, such as inputting data into the CDU (control display unit), and monitoring the EICAS (engine indicating and crew alerting system) for specific alerts.

## CPM Development

Based on the HTAs, the CPM-GOMS model were developed to represent all pre-flight and monitoring phase actions. For this analysis, we also recruited three engineers from Northrop Grumman Corporation to perform flight tasks using the UH-60V flight simulator. Each engineer had a minimum of 10 years of experience in helicopter operations and simulation training. Their backgrounds included military training, aviation safety, and human factors engineering. All three engineers had extensive experience with the specific UH-60V simulator. Each engineer provided informed consent for the experimental study. They subsequently performed simulator test trials involving each phase of flight. Some trials also required emergency procedures. In total, data was recorded for 21 trials. This included the flight tasks that were performed for certain periods of time and cockpit button press logs. The logs also indicated if a press was made in error. Video feeds were also captured with no audio. De-identified data was shared with the research team for the CPM development.

Standard interface operation times were taken from the literature (Olson & Olson, 1995) and used to code most of the flight simulation operations as part of the CPMs and to determine average task completion times. A requirement for custom task operators was identified from the cockpit button press logs and video feeds of the simulation trials. Two custom task operators were created to code specific piloting behaviors in Cogulator, including flipping a checklist page (a motor operation) and comparing checklist items (a cognitive operation). Actual action times were recorded for the three participating engineers to estimate mean expert error-free task operation times in pre-flight and flight tasks. The CPMs were subsequently run in Cogulator, including all task perceptual, cognitive, and motor operations to estimate overall flight task time (Estes, 2017).

In addition to task time, the CPMs also allowed for estimate of working memory (WM) demands. Each time specific GOMSL task operators are coded with Cogulator (i.e., Recall, Look, Search, Perceptual Processor, Listen or Think) with reference to a named item to be manipulated in WM, Cogulator counts an active WM "chunk". Chunks are cohesive "pieces of information" that need to be "kept in mind" by system users to perform specific tasks. Black Hawk pilots need to maintain many chunks in WM for pre-flight and monitoring tasks. In the case of a checklist review, the three most recent task steps may be chunks in WM. Related to this, prior cognitive science research (Gobet & Clarkson, 2004), has indicated that holding four chunks of information in WM during task performance should be considered as a threshold of high workload. Cogulator provides WM counts in conjunction with each task operation and can be used to estimate the average chunk count per task time (Figure 3).

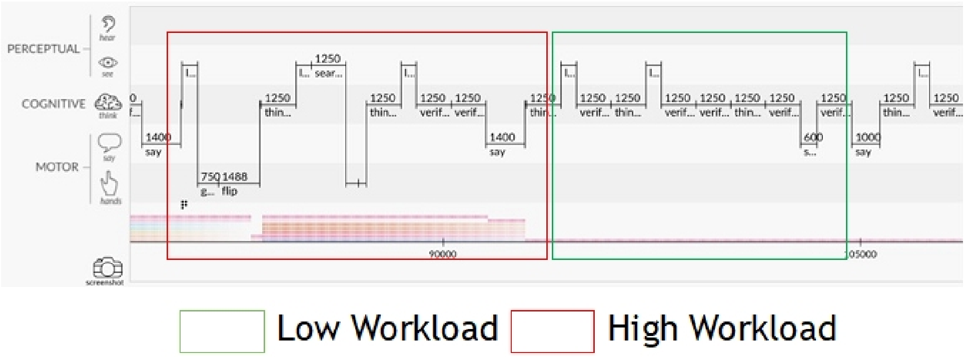


**Figure 3:** Cogulator representation of task operations requiring perceptual, cognitive, and motor processor resources. Memory chunk counts occurring at each operation are represented by layered graph at bottom of Gantt chart. Average chunk count per time is calculated for complete task. Task operator time estimates are presented in milliseconds (e.g. "flip" at 1488ms). Chart shows last portion of monitoring phase (Emergency Task 1) as posing high workload and Emergency Task 2 representing a relatively low workload sequence.

## Expert Interviews

Following the coding of the CPMs using Cogulator, expert interviews were conducted to validate model content. Two experts were recruited to the study with one specializing in the flight operations protocol and procedures, and the other specializing in UH-60 helicopter design and operation. In the interviews, the experts were posed with the following questions: 1) how accurate do the models represent the pre-flight and monitoring phases; 2) how accurate is the model estimate of the number of checklist reviews during either pre-flight or monitoring tasks; 3) which tasks in either pre-flight or monitoring pose high pilot workload; 4) reflecting on past flight task performance, how effectively do the models reveal cognitive challenges associated with emergency procedures under high workload and time constraints; and 5) what changes to flight procedures or checklist items could reduce pilot workload. The interviews were structured and lasted approximately 60 minutes each. Interviews were recorded for subsequent analysis.

# RESULTS

## Model Outcomes

The CPMs, as executed with Cogulator, provided task time estimates, working memory demand estimates, and subjective workload assessments for the pre-flight tasks (Table 3) and monitoring tasks (Table 4). The pre-flight model identified three tasks (3, 4 and 7; see Table 3) to pose high pilot workload. As for the monitoring tasks, an additional three emergency tasks (Emer-1, Emer-4 and Emer-5) were found to pose high workload. The model estimate of total pre-flight task time was compared with observed data from the three pilots. The model output had a total time of 4 minutes and 31 seconds, and the observed average time (from 21 trials) was 4 minutes and 32 seconds (for error-free completion). Since the monitoring tasks were performed in a loop, and emergency tasks were not always present in the 21 trials, observed task completion times were not available for this analysis. Video feeds were used to identify how often checklist reviews occurred, and this was incorporated in the CPM-GOMS models.

**Table 3.** Task time estimates, WM demands, and workload estimates for pre-flight tasks. Working memory demand is average chunk count over course of pre-flight phase or for specific task. Four or more chunks is considered threshold for high overall workload (Gobet & Clarkson, 2004). Bold rows represent higher workload tasks.

| Flight Phase | Task | Task Time Estimate | Type of Operations | Working Memory Demand | Subjective Workload |
|---|---|---|---|---|---|
| ***Pre-flight*** | ***All*** | ***271.7s*** | ***All operations are listed below*** | ***4.3*** | ***High*** |
| Pre-flight | 1 | 7.2s | Perceptual (look, search); Cognitive (think, store, ignore); Motor (keystroke, hands) | 1.5 | Low |
| Pre-flight | 2 | 13.1s | Perceptual (look, search); Cognitive (think, store, verify, ignore); Motor (keystroke, hands, say) | 0.9 | Low |

(Continued)

**Table 3.** Continued

| Flight Phase | Task | Task Time Estimate | Type of Operations | Working Memory Demand | Subjective Workload |
|---|---|---|---|---|---|
| **Pre-flight** | **3** | **57s** | **Perceptual (look, search, read); Cognitive (think, store, verify, recall, ignore); Motor (keystroke, hands, say, type)** | **4.8** | **High** |
| **Pre-flight** | **4** | **54.4s** | **Perceptual (look, search); Cognitive (think, store, verify, recall, ignore); Motor (keystroke, hands, say, type)** | **6.4** | **High** |
| Pre-flight | 5 | 6.2s | Perceptual (look, search); Cognitive (think, store, verify, recall, ignore) | 0.7 | Low |
| Pre-flight | 6 | 9.4s | Perceptual (look, search); Cognitive (think, store, verify, recall, ignore) | 0.8 | Low |
| **Pre-flight** | **7** | **116s** | **Perceptual (look, search); Cognitive (think, store, verify, recall, ignore); Motor (keystroke, hands, say, turn, type)** | **4.4** | **High** |
| Pre-flight | 8 | 4.9s | Perceptual (look); Cognitive (verify); Motor (hands, turn) | 0 | 0 |
| Pre-flight | 9 | 3.5s | Cognitive (verify, think); Motor (say) | 0 | 0 |

**Table 4.** Task time estimates, WM demands, and workload estimates for flight monitoring tasks. Average chunk counts are shown for phase of flight and specific tasks. Four or more chunks is considered threshold for high overall workload (Gobet & Clarkson, 2004). Bold rows represent higher workload tasks.

| Flight Phase | Task | Task Time Estimate | Type of Operations | Working Memory Demand | Subjective workload |
|---|---|---|---|---|---|
| ***Monitoring*** | ***All*** | ***302.8s*** | ***All operations are listed below*** | ***4.9*** | ***High*** |
| Monitoring | 1 | 15.2s | Perceptual (look, search); Cognitive (think, verify); Motor (keystroke, hands, say) | 0.3 | Low |
| Monitoring | 2 | 18.7s | Perceptual (look, search); Cognitive (think, verify, store) | 2.4 | Low |
| **Monitoring** | **Emer-1** | **63.2s** | **Perceptual (look, search); Cognitive (think, store, verify, recall, ignore, compare); Motor (keystroke, hands, say, grasp, flip)** | **6** | **High** |
| Monitoring | Emer-2 | 9.7s | Perceptual (look); Cognitive (think, verify); Motor (say) | 0 | Low |
| Monitoring | Emer-3 | 20.1s | Perceptual (look, search); Cognitive (think, verify); Motor (keystroke, hands, say) | 1.6 | Low |
| **Monitoring** | **Emer-4** | **116.1s** | **Perceptual (look); Cognitive (think, store, verify, recall, ignore); Motor (keystroke, hands, say, click, flip)** | **5.3** | **High** |
| **Monitoring** | **Emer-5** | **47.8s** | **Perceptual (look, search); Cognitive (think, store, verify, recall, ignore); Motor (keystroke, hands, say, type, flip, turn)** | **5.1** | **High** |
| Monitoring | Emer-6 | 12.1s | Perceptual (look); Cognitive (think, verify); Motor (say) | 0 | 0 |

## Interview Feedback

The expert interviews revealed the model content to be perceived as accurate relative to actual pre-flight and monitoring task performance. Both experts agreed with the sequence and representation of operations in the CPM-GOMS models. When asked about checklist review frequency, the experts agreed with the estimated number of reviews embedded in the CPMs.

Regarding identification of high workload tasks, there was some variation in expert opinions but general agreement with our model-based observations. Expert 1 identified pre-flight Task 4 to be most cognitively demanding along with Emer(gency) Task 1 (the pitch miscompare) and Emer-5 (engine emergency) to be the most cognitively challenging monitoring tasks. Expert 2 identified pre-flight Task 4 to be most cognitively demanding and Emer-4 (pop-up weather event) to be the most cognitively challenging monitoring task. In past Black Hawk incidents (Aviation Safety Network, 2024), Emer-5 (engine failure) and Emer-1 (pitch miscompare) were often misidentified by pilots. The study experts also mentioned that inputting data in pre-flight Tasks 4 and 7 and monitoring Emer-4 task were activities that might elevate cognitive workload due to the number of pieces of information that need to be maintained in working memory.

## DISCUSSION

The model outcomes for pre-flight and monitoring tasks were found to align closely with observed expert performance in terms of task completion time. On this basis, the models appeared to be valid representations of the piloting tasks, underscoring the utility of such methods for predicting task performance.

The expert interviews yielded critical feedback that further validated the CPM content and reinforced our model-based cognitive workload estimates for various sub-tasks. The experts agreed with the model representation of the sequence of tasks and subtasks as well as the operational content of tasks. The experts offered that the modeled cognitive processes and task sequences were accurate reflections of real-world performance. Experts highlighted specific emergency tasks as being particularly challenging due to high cognitive workload. These tasks were a subset of those identified by the CPMs as posing high WM demands. For these tasks, pilots may struggle to complete all necessary checks within time constraints imposed by the emergency condition. In such situations, there is a need for pilots to prioritize the most urgent tasks while working to identify and resolve the primary flight issue. This combination of cognitive demands can significantly increase workload. The expert feedback further validated the CPM methodology as a basis for identifying areas of high cognitive workload as a basis for re-designing procedures and interfaces to support pilot information processing and performance in high-pressure situations (Amalberti & Wioland, 2020).

On the basis of these results, we proposed several design recommendations towards reducing the cognitive workload experienced by pilots during flight operations. The recommendations focus on adaptive aiding solutions that could support pilots in high-workload situations. Such aiding has shown promise in related military aviation research (Aguilar Reyes et al., 2023). The proposed recommendations also seek to enhance pilot capability to process information and make critical decisions efficiently:

- Auto-Text Completion: To respond to high cognitive workload during data entry (such as in Emer-4), we recommend auto-text completion for entering RF IDs, and crew and weight information. This feature may also allow pilots to focus on higher-priority tasks by minimizing time and effort required for routine entries.
- Flight Management System (FMS) Prompts for Re-Confirming Flight Plan (FPLN) Entries: For flight plan selection and reconfiguration (during pre-flight Task 7 and Emer-4), we recommend providing prompts in the FMS to re-confirm FPLN entries. Not only may such prompts reduce pilot WM demands they can help reduce the likelihood of task errors during critical phases of flight. The prompts my also promote task efficiency as they confirm for pilots that all necessary information has been input correctly without need to revisit previous entries (cf., Wickens, 2002).
- Automated Alerts for Out-of-Range Parameter Settings: Automated alerts on parameters, such as heading settings (in the case of Emer-1), can serve to immediately notify pilots of discrepancies of flight from FPLN, allowing them to make quick corrections. While alerts exist for automated systems, additional alerts should be implemented when sensor data from different sources do not match, which can be the case in a pitch miscompare emergency. Alerts can reduce piloting monitoring workload and promote task efficiency
- Drop-Down List of Nearest Accessible Landing Zones (LZs): In emergency situations (such as Emer-4 or Emer-5), providing a drop-down list of the nearest accessible landing zones can expedite the decision-making process for re-routing. This feature recommendation helps pilots quickly identify the safest and most feasible landing options without extensive manual searches; thereby, reducing workload and increasing operational efficiency.
- Temporary Shutdown of Non-Relevant Displays: During emergencies, temporarily shut down displays that are not relevant to the primary flight task (e.g., ETA analysis) can help pilots concentrate on the most critical information. This reduction in display may reduce pilot visual workload, and improve situation awareness and decision-making efficiency.

Returning to the interview outcomes, the experts offered that any flight task redesigns need to consider how simplifications in checklist items could in turn increase memory load for pilots. That is, reducing "knowledge in the world" could translate to a greater requirement for "knowledge in the head" (Norman, 2013). In general, the experts were of the opinion that task redesigns could be beneficial for pilot performance and flight safety.

### Implications for Future Research and Practice

Our study methodology and findings highlight the importance of incorporating CPM in the evaluation and design of piloting procedures. By leveraging tools like CPM-GOMS and Cogulator, aviation professionals can identify and address cognitively demanding tasks early in the system design process to support effective and efficient interface design. Future research should further explore the application of CPMs for analysing WM demands and cognitive workload for various phases of flight and specific flight operations. Additionally, expanding the current set of CPM task operators and automating certain aspects of model coding could extend the implementation of CPM methodology to other broader contexts and complex systems also requiring the use of SOPs and checklists.

## CONCLUSION

The human factors evaluation method employed in this study, combining HTA, CPM-GOMS modeling, and expert interviews, provided a robust framework for assessing and refining the cognitive demands of pre-flight and flight tasks in the UH-60V simulator. The insights gained from this research may contribute to the development of cognitively efficient checklists and flight procedures, ultimately enhancing pilot performance and aircraft safety.

## ACKNOWLEDGMENT

This research was partially funded by the U.S. Defence Advanced Research Projects Agency under contract #HR00112220004. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies. The authors would like to acknowledge Margarete Groll for providing expert opinions.

## REFERENCES

Aguilar Reyes, C. I., Wozniak, D., Ham, A., & Zahabi, M. (2023). Design and evaluation of an adaptive virtual reality training system. Virtual Reality, 27(3), 2509–2528.

Amalberti, R., & Wioland, L. (2020). Human error in aviation. In Aviation safety, human factors-system engineering-flight operations-economics-strategies-management (pp. 91–108). CRC Press.

Aviation Safety Network. (2024). Aviation Safety Network - ASN safety database : UH-60 search.

Card, S. K. (2018). The psychology of human-computer interaction. Crc Press.

Estes, S. (2017). Cogulator: A cognitive modeling calculator.

Fujizawa, B. T., Tischler, M. B., Ott, C. R., & Blanken, C. L. (2014). UH-60 partial authority modernized control laws for improved handling qualities in the degraded visual environment. American Helicopter Society 70th Annual Forum.

Gawande, A. (2010). Checklist manifesto, the (HB). Penguin Books India.

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four… or is it two? Memory, 12(6), 732–747.

Goodwin, M. (2017). A Training Effectiveness Evaluation of UH-60A/L Simulated Environments: An Interdisciplinary Approach.

Havir, T. J., Durbin, D. B., Frederick, L. J., & Hicks, J. S. (2006). Human factors assessment of the UH-60M crew station during the limited user test (LUT). Aberdeen Proving Ground (MD): Army Research Laboratory (US).

Jones, D. G., & Endsley, M. R. (1996). Sources of situation awareness errors in aviation. Aviation, Space, and Environmental Medicine, 67(6), 507–512.

Park, J., & Zahabi, M. (2024). A review of human performance models for prediction of driver behavior and interactions with in-vehicle technology. Human Factors, 66(4), 1249–1275.

Norman, D. A. (2013). The Design of Everyday Things. New York: Basic Books (Perseus Books).

Thornton, R. C., Kaempf, G. L., Zeller, J. L., & McAnulty, D. M. (1992). An evaluation of crew coordination and performance during a simulated UH-60 helicopter mission. Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science, 3(2), 159–177.

Wiegmann, D. A., & Shappell, S. A. (2001). Human error perspectives in aviation. The International Journal of Aviation Psychology, 11(4), 341–357.