Human-Centered Design and User Experience, Vol. 114, 2023, 11–20
https://doi.org/10.54941/ahfe1004216

A Human-Centered Approach to Artificial Intelligence Applications in Naval Aviation

Mitchell J. Tindall[1], Beth F. Wheeler Atkinson[1], Jordan M. Sanders[2], Sarah C. Beadle[2], and James A. Pharmer[1]

[1]Naval Air Warfare Center Training Systems Division, Orlando, FL 32826, USA
[2]Naval Research Enterprise Internship Program, Orlando, FL 32826, USA
[3]Naval Air Warfare Center Aircraft Division, Patuxent River, MD 20670, USA

ABSTRACT

The doubling of artificial intelligence (AI) performance every six months (Sevilla et al., 2022) during the last decade necessitates that the application of these capabilities in high stakes settings not be done arbitrarily. Defining a structured, human-centered process increases the likelihood that the application of AI is done safely, effectively, and efficiently. Such a process, which considers both AI and automation, should start by identifying clear definitions to guide categorization of capabilities. A recent literature review identified 28 definitions for AI (Collins et al., 2021), to include AI being "…the ability of a machine to perform cognitive functions…problem solving, [and] decision-making." Due to their complexity, investments in developing models can reach the millions (Maslej et al., 2023). Alternatively, automation can be defined as "…something which runs itself with little to no human interaction…" and guided by specific rules (GeeksforGeeks, 2022). Unique to AI is the ability to learn and evolve (GeeksforGeeks, 2022). With these definitions, the next step should focus on a comprehensive review of targeted domain tasks. This would include understanding the associated knowledge, skills, and abilities (KSAs), as well as the tasks' criticality, frequency, and difficulty. Such information is generally a product of (cognitive) task analysis and/or front end analysis and is valuable when building criteria for the appropriateness of AI/automation. A recommended next step should include reengagement with experienced end-users, which is imperative for ensuring a comprehensive understanding of tasks and for yielding valuable insight into AI applications. This poster will provide an overview of the steps undertaken for initial consideration of AI and automation within a Navy domain, to include exclusion criteria and lessons learned with regard to applying this process. Finally, results will include estimated applicability of AI/automation technologies as related to current tasking in relevant aviation platforms.

Keywords: Artificial intelligence, Naval aviation, Human-centered

INTRODUCTION

The rapid pace of technology improvements and developments is a concept that affects individuals in every sector of life, from personal devices to enhancing job performance. However, emerging technological advances in component technologies such as artificial intelligence continues to increase the rate at which innovative solutions are available. In fact, Sevilla and colleages (2022) found that over the last decade the performance of artificial intelligence (AI) systems has doubled every six months, greatly outperforming Moore's Law. Considering this pace of advancement, application of these capabilities in high stakes settings like military domains should not be done arbitrarily nor haphazardly. Further, part of a user centered identification of reasonable applications and requirements for AI technology should consider not only the technological capabilities but also the human needs and abilities to effectively rely on AI implementations. While literature on human-autonomy teaming has documented effective approaches to the application of AI and automation (O'Neill, McNeese, Barron, & Schelble, 2020; Huang, Cooke, Johnson, Lematta, Bhatti, Barnes, & Holder, 2020) the approach outlined in this paper focuses on mission specific tasks and adapting/integrating with the warfighter in their context, vice approaches to design AI well before it reaches the end user.

In military domains, a standardized process during early phases of acquisition programs exists for defining system requirements and appropriate technological solutions. However, this process does not specifically take into account fast evolving capabilities like AI to assist the human operator. In an environment that encourages *speed to the fleet transitions* and *fail faster* technology investigations, AI offers promising opportunities. As such, the Chief of Naval Operations' 2022 Navigation Plan implores the Naval Forces to: "Leverage [artificial intelligence] to support … warfighting… by 2023, launch a framework to identify gaps and accelerate delivery of AI-enabled capabilities to the Fleet and Navy enterprises."

To support these calls for AI technology adoption, with emphasis on maximizing investments, what is necessary is a systematic, human-centered approach to ensure the application of this technology is done safely, effectively, and in a way that ensures optimal return on investment (ROI). As a rapidly evolving technology, the optimal applications for AI within high stakes, complex systems like naval aviation offer unique use cases that may translate to commercial applications in the future. This paper outlines a proposed process for understanding and defining potential insertion points for automation and AI technologies that sets operational definitions for organization, standardizes an objective method that leverages existing documentation and subject matter expertise, and maintains a human-centered approach to requirements and design.

DEFINING CONCEPTS

For the purposes of this effort, primary concepts for consideration were *AI* and *Automation*. Generally speaking, *Automation* refers to technology used to perform tasks or processes without direct involvement from humans, functioning independently to reduce the need for constant human intervention. Alternatively, *AI* refers to technological solutions that can perform tasks that typically require human intelligence (e.g., learning, decision-making, problem-solving) by leveraging algorithms and models that enable functions that are analogous to human cognition. However, due to the rapid evolution of this technology in recent years and the variety of solutions within this Automation-to-AI spectrum, there are a plethora of ways to define these concepts.

Collins, et al. (2021) found 28 definitions for AI in their systematic literature review. As an example, one of these definitions was AI, "…is defined as the ability of a machine to perform cognitive functions that we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem solving, decision-making, and even demonstrating creativity." Additionally, AI models have been known to cost thousands and even millions of dollars (PaLM, a "…large language model launched in 2022….") (Maslej, N. et al., 2023). Automation, on the other hand, is defined "as the execution by a machine agent (usually a computer) of a function that was previously carried out by a human" (Parasuraman & Riley, 1997). There are several key differences between AI and automation, including that "AI involves learning and evolving," while automation does not (GeeksforGeeks, 2022; see Table 1). Understanding the state of the art and practice of AI and automation is an important first step in the development of a systematic approach to making AI/automation decisions. This step should result in verbiage that aligns with AI/automation's current functionality (e.g., analysis, verification, synthesis, aggregation).

Table 1. Comparison of automation and artificial intelligence definitions with examples.

Key Concept	Automation	Artificial Intelligence
Functionality	Repetitive tasks[1]; automated processes[2]	Complex tasks, decision-making involved[3]
Human Involvement	Limited	Minimal to none
Learning	Follows pre-set instructions[1]	Learns from data[1]; improved performance[3]
Adaptability	Limited in new situations[1]	Adapt or evolve to changing or new scenarios[1]
Example Use Cases	Assembly lines, e-mail filter, coffee-makers, customer support[1]	Self-driving cars, speech recognition, machine learning[1]

[1]GeeksforGeeks, 2022; [2]Parasuraman & Riley, 1997; [3]Collins et al., 2021

SCOPING THE ANALYSIS

Within domains such as military training, an important early step in the process is providing a valid and scoped use case. Due to the inherent complexity and variety of systems and capabilities within the military, this process helps manage expectations, minimize scope creep, and eases identification of relevant documentation and subject matter experts (SMEs). To start, the pre-requisite questions utilized were:

- What mission do I want the AI to support?
- Who in that mission do I want to focus on?

- What information security classification do I want to maintain?
- What platforms support the mission I am focusing on?
- Am I able to obtain Front End Analysis tasking data on this mission in these platforms?
- What phase of the acquisition lifecycle is the system in and is there potential funding to implement a change in the future?

DATA ANALYSIS PROCESS

The proposed data analysis process is an iterative multi-step process intended to leverage traditional training system analysis documentation and SME input to provide a comprehensive evaluation of technology opportunities. Figure 1 provides an overview of the primary steps identified during a feasibility analysis conducted within a naval aviation domain.

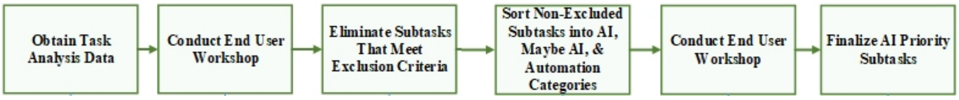

Figure 1: Human-centered artificial intelligence data analysis process.

Step 1. Obtain Task Analysis Data

With the scope outlined by the pre-requisite questions, the next step is to contact relevant stakeholders to gain access to relevant documentation. Types of documentation might include task analysis data, interface design documentation, software user manuals, training material, operational manuals, tactical procedure documentation and the like. Critical aspects of those documents include a list of tasks & sub tasks, specific steps for performing tasks, the knowledge, skills and abilities required to perform tasks, the criticality, difficulty and frequency of task performance, information regarding how a graphic user interface (GUI) is used to perform tasks and the context of task performance. Together this information provides an excellent starting point for further scoping AI development initiatives for communities, platforms and capabilities.

Specifically, these pieces of information are necessary for building criteria for the appropriateness of AI/automation for performing tasks. For example, tasks that contain verbs such as analysis, verification, synthesis, etc..., may be well-suited for AI/automation given the current state of the technology. This narrowing of the task list is crucial before the next steps when SMEs are engaged. To operate platforms and perform missions in military contexts, operators can perform hundreds of tasks and thousands of steps. Engaging SMEs with task, mission and domain information with thousands of data points would be inefficient and unproductive. Therefore, a scoped list of tasks that qualify as good candidates for AI or automation should facilitate highly productive SME engagements.

Step 2. Conduct End User Workshops

While task analysis is a useful starting point for narrowing the scope of an AI/automation development effort, engagement with experienced end-users is imperative for ensuring a detailed and comprehensive understanding of tasks and the job. Additionally, these engagements may yield valuable insight into where to best insert AI into a job that cannot be derived from task analysis data.

- Setup meetings with end users
- Elaborate on task analysis data
- Establish initial end user ideas about AI

End user workshops are particularly beneficial for learning additional mission context that may impact an AI application or expand the scope of a mission. For example, there may be a known gap in sensor performance or an external variable like weather that adds complexity to a given task. Often, these components are not included in Task Analysis data, but would have implications for an AI-based solution. Additionally, by talking to groups of end users, you are able to examine where training is used to supplement complex tasking. Other insights derived from end users that cannot be gleaned from Task Analysis are identifying certain tasks where there is variability in human performance, particularly between novice and expert users, that can highlight a lengthy time-to-train or need for a decision aid.

End user workshops also serve an important role in developing the appropriate language and mission understanding for the human factors team. Identifying the sequence of events and discussing the goals of the mission are critical prior to examining the task data or observing users performing the mission. The secondary benefit of engaging end users early in the process is buy-in and shared interest in the effort as they return to supporting the task. By prompting these stakeholders early in the process with task-related questions and engaging them throughout the lifecycle they share investment in the task. This partnership is critical for their role as gatekeepers into a community of experts and yields ongoing conversation on appropriate tasks for consideration.

Step 3. Eliminate Subtasks via Exclusion Criteria

As previously stated, there is an abundance of potential tasks to examine in any military mission context. At this stage, heuristics are developed to further reduce the amount of potential subtasks to consider. The focus of task reduction is to examine where an AI solution would not be appropriate-both in terms of mission difficulty/criticality and return on investment for a technological solution. Within this review, exclusion criteria to consider include:

- Does not apply to desired mission
- All steps are critical
- All steps are not difficult
- Less frequent than once every 6 weeks

- Pertain to non-priority positions / roles
- Require critical thinking skills

The first step to scope is removing tasks that do not support the mission of your focus. Due to the high risk nature of military contexts, any task with subtasks that are deemed 'critical' are not considered- this again impacts our security classification and the overall risk of an AI solution. Tasks that are rarely performed, not difficult, or non-priority are removed from consideration as there is little impact to mission performance with an AI system being added. Last, understanding what knowledge, skills and abilities are necessary for performing tasks can help determine if and what type of AI could be leveraged to perform that tasking. However, KSA information is not always available or descriptive enough for making these determinations. When KSA data lacks sufficient detail, often an analysis of subtasks by psychologists illuminates whether things like critical thinking and decision making are necessary for subtasks like, analyze, determine, verify, detect, identify, monitor, etc.

For example, the review of a checklist required prior to flight is critical to safety of flight assessments to determine if an aircraft can meet go/no go criteria. While potentially a fit for an assistive automation process, taking the human out of the loop in this situation may have dire consequences.

Other criteria for the elimination of tasks can include:

- Communication
- Requires a human
- Performing a check
- Performing a set of procedures
- Minimal decision-making
- Starting / setting up a system
- Outside designated classification
- Utilizing an existing application with no obvious AI application
- Subtask goal does not align with role / position

These considerations assist with ensuring an AI solution has impact to support the operator versus take over their role on the mission.

Step 4. Sort Non-Excluded Subtasks Into Categories

This step focuses on organizing remaining subtasks within one of three categories: AI, Maybe AI, or Automation. Subtasks that require decision-making with several steps sort within the *AI* category. Alternatively, simplistic decision-making subtasks move within the *Automation* category. The remaining subtasks, organized as *Maybe AI*, are likely subtasks that are somewhat ambiguous in wording or due to limited domain context are not easily organized in one of the former categories. For example, subtasks that involve decision-making but lack details to determine the complexity of associated steps may require additional engagement with end users to determine if they better align with *AI* or *Automation.*

Step 5. Conduct End User Workshops

End user workshops at this phase in the process are targeted on verifying the scoping from the previous step, refining an understanding of ambiguously described tasks and subtasks for further scoping, and offering additional insight into potential recommendations for technology solutions not afforded from task analysis data. As such this step includes:

- Fill in blanks of task analysis data.
- Attain classification verification (i.e., *AI* or *Automation*) on subtasks.
- Validate a final list of subtasks.

Step 6. Finalize AI Priority Subtasks

During this final step prior to prototype investment and development a final narrowed done list of tasks, sub tasks and steps should be complete. A workshop that should include all relevant stakeholders (e.g., scientists, computer scientists, fleet end users, software engineers, program managers, leadership) will rate the tasks based on several criteria. That criteria includes prioritization, AI/automation type and vulnerability/exploitability. Depending on the size of the final list, ranking ordering or simple high, medium or low priority rating could be used to determine where initial prototype investment should focus. There are roughly seven types of AI (e.g., theory of mind, natural language processing, neural networks) (Joshi, 2019). To start envisioning the architecture of a prototype, the workshop group should determine what type of AI/automation is best suited to performing tasks. This step helps determine cost of development, an important consideration for maximizing ROI and potentially revisiting prioritization rankings/ratings. Unique to military contexts is the fact that there are forces motivated to neutralize any capability advantage you maintain. While recent advances in AI have proliferated at an exponential rate, there are still notable limitations to each type of AI. Those limitations could result in vulnerabilities that adversaries will try to exploit. As such, the workshop group should consider several facets of vulnerability and exploitability of the AI type selected from the previous step. These facets should include whether the AI and a human are equally easily exploited, whether the AI but not a human would be easily exploited, whether the AI has limited chance of being exploited and whether the AI has no chance of being exploited. The vulnerability/exploitability criteria not only offers an opportunity to further refine your priorities list, it ensures investments are not made in a system that could result in a significant vulnerability that inhibits mission performance or at worst puts lives at risk.

PROCESS IN PRACTICE

This process was implemented for two aviation platforms, scoped to a single mission set. The results provide a preliminary look at how this process might assist decision makers with scoping initial discussion for emerging AI technology to maximize resources.

Within the first aviation platform, an existing front end analysis provided a total of 1,783 relevant subtasks or tasks that contained no subtasks. In

Step 3 of the outlined process, 1,670 subtasks were excluded and categorized in a *None* category for relevance to AI or Automation. During Step 4 of the process, the remaining subtasks were organized in the remaining three categories. Figure 2 provides a summary of the results with additional data on the justification for categorization.

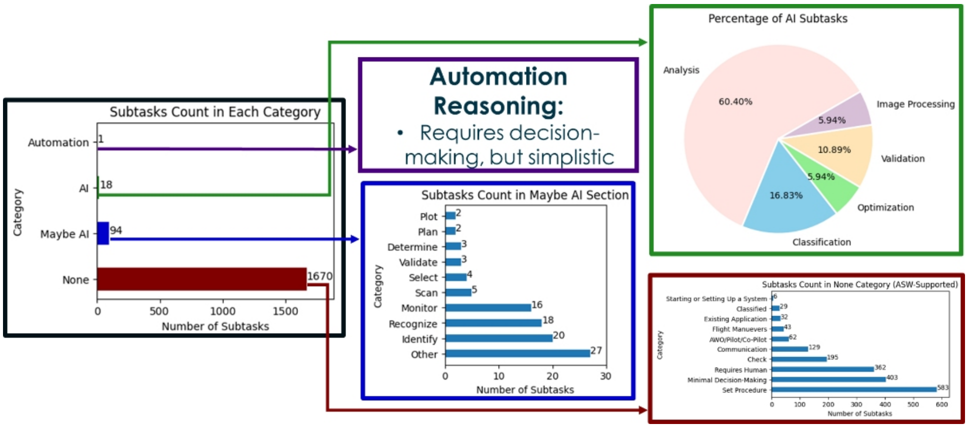

Figure 2: Example analysis of AI and automation from step 4 of proposed process.

While there were fewer available subtasks within the second aviation platform documentation, a total of 234 relevant subtasks or tasks that contained no subtasks were identified. Exclusion criteria (Step 3) resulted in 157 subtasks being categorized as *None*. However, due to the robust subtask data to include criticality and prioritization information available for this data set, Step 4 analyses included not only categorization of subtasks as *AI* or *Automation*, but also the types of technology that might be beneficial. Figure 3 provides a summary of the results with additional data on the justification for categorization.

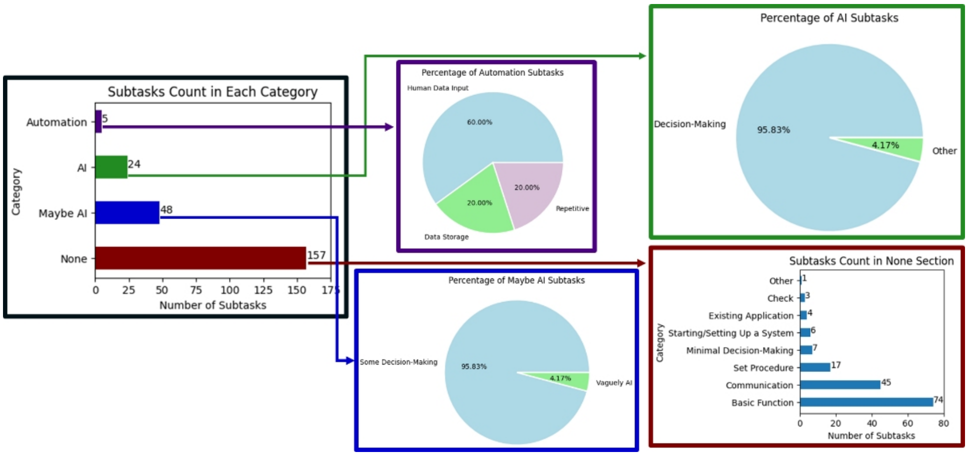

Figure 3: Example analysis of AI and automation from step 4 of proposed process.

CONCLUSION

Defining the current state of the technology, leveraging systematically collected task analysis data, and frequently engaging experienced end users will increase the likelihood of the safe and effective application of AI/ automation development efforts. In addition, such a systematic approach to making these decisions enhances decisions regarding development and implementation with data to support tradeoffs and prioritization. For these reasons, there is an increased likelihood that solutions will maximize organizations' return on investment and benefits associated with AI or Automation technologies.

While this process has been used to guide preliminary findings within two aviation platforms for a specific mission set, the resulting human centered system for making AI determinations is expected to be generalizable across domains or uses cases. As more communities and platforms within naval aviation or other domains attempt to adopt AI, we propose that this system would provide a standardized guide for maximizing AI implementation.

It is important to note that targeted AI implementation within a use case will likely maximize benefits to the user community and organizations; however, there are other barriers to adoption of technology that must be considered. For example, recent research with radiologists highlighted that while AI technology for human operator assistance offers useful benefits, "biases in humans' use of AI assistance eliminate these potential gains" (Agarwal, Moehring, Rajpurkar, & Salz, 2023). That is, results of this study suggest that policies that encourage human users to "work next to as opposed to with AI" provided optimal results (Agarwal et al., 2023).

As this effort continues, consistent emphasis on iterative end user engagement will be sought. These working groups to seek end user feedback will afford additional contextualized perspective on the design and implementation of AI solutions. Further, as needs are refined and prioritized, end user engagement is intended to increase buy-in to facilitate effective transition. Aspects of these future workshops will focus on known challenges associated with transparency, trust calibration, situation awareness, workload balancing, vulnerability & exploitability, as well as considerations for policy for implementation to maximize benefits when AI technology is fielded.

ACKNOWLEDGMENT

The views of the authors expressed herein do not necessarily represent those of the U.S. Navy or Department of Defense (DoD). Presentation of this material does not constitute or imply its endorsement, recommendation, or favoring by the DoD. NAWCTSD Public Release 23-ORLXXX Distribution Statement A – Approved for public release; distribution is unlimited.

REFERENCES

Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2023). *Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology* (No. w31422). National Bureau of Economic Research.

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*. https://doi.org/10.1016/j.ijinfomgt.2021.102383

GeeksforGeeks. (2022, July 19). *Difference between artificial intelligence and automation*. https://www.geeksforgeeks.org/difference-between-artificial-intelligence-and-automation/

Huang, L., Cooke, N., Johnson, C., Lematta, G., Bhatti, S., Barnes, M., & Holder, E. (2020). Human-Autonomy Teaming: Interaction Metrics and Models for Next Generation Combat Vehicle Concepts.

Joshi, N. (2019). 7 types of artificial intelligence. *Forbes*. https://www.forbes.com/sites/cognitiveworld/2019/06/19/7-types-of-artificial-intelligence/?sh=3fce3de6233e

Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Parli, V., Shoham, Y., Wald, R., Clark, J., & Perrault, R. (2023). *The AI Index 2023 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, California. https://aiindex.stanford.edu/report/

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2022). Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors*, *64*(5), 904–938.

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230–253.

Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M., & Villalobos, P. (2022, July). Compute trends across three eras of machine learning. In *2022 International Joint Conference on Neural Network (IJCNN)* (pp. 1–8). IEEE.