SYM-AM-25-368

# Excerpt from the Proceedings of the Twenty-Second Annual Acquisition Research Symposium and Innovation Summit

## Volume III

## Augmenting Intelligence: Acquiring Trustworthy Technology

**Published: May 5, 2025**

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).


ACQUISITION RESEARCH PROGRAM
DEPARTMENT OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL

# Augmenting Intelligence: Acquiring Trustworthy Technology

**Carol J. Smith—**is a Principal Research Scientist at the CMU Software Engineering Institute in the AI Division. Smith has been conducting research to improve experiences with complex systems for over 20 years. Smith trailblazed human-computer interaction (HCI) research for AI, autonomous vehicles, and other emerging technologies and is recognized globally as an advocate for human-centered AI. Smith is a co-author of the DoD Defense Innovation Unit Responsible AI Guidelines and has presented over 250 talks and workshops around the world. She teaches at CMU and holds an MS in HCI from DePaul University. [cjsmith@sei.cmu.edu]

### Abstract

Procuring understandable systems that improve human effectiveness and reduce (or at least maintain) overall risk has become even more complex when considering the acquisition of artificial intelligence (AI) systems. Current acquisition guidance provides a strong foundation but is not sufficient to identify and effectively assess emerging technologies such as generative artificial intelligence. These systems can provide value and be exceedingly helpful in the right situation. This paper provides actionable guidance to Navy acquisition teams so that they can quickly and effectively identify and procure the best AI systems and reduce risks associated with these major investments.

## Introduction

This paper provides guidance for acquisition teams to quickly select, test, and implement AI systems with a streamlined practice that informs procurement teams, enables operational experimentation, and ensures the continuity of capabilities. Successful acquisition and adoption of emerging technologies requires an ability to identify aspects that will enable their trustworthiness. This work builds on the Software Acquisition Pathway (Defense Acquisition University [DAU], 2020) which the recent Department of Defense (DoD) memo (Hegseth, 2025) directs all DoD components to adopt. In addition, the paper incorporates decades of research and experience developing and designing complex and dynamic systems that work with, and for, humans.

The approaches described in this paper will enable the United States to gain an early advantage by quickly identifying the best AI solutions to achieve our goals, that meet the needs of the workforce, and that reduce cost and risk. The introduction of generative AI systems has ignited a significant leap in awareness of the capabilities of AI and will be specifically addressed in the paper. This guidance is designed for use cases that involve some risk (forecasting, planning, anomaly detection, etc.). This guidance can also be used for very low risk systems (e.g., movie recommenders), but it is not appropriate for extremely high-risk systems such as robotics or weapons systems. This guidance assumes that the organization is making a significant investment in the new system and is interested in assuring it is adopted effectively.

## Informed Approach

Successful acquisition starts with a brief assessment to determine the organizations' readiness for AI technology and that AI is a match with their needs. While there are many formal methodologies for requirements gathering, this approach is focused on gaining relevant knowledge and informing the acquisition process. There are two areas of focus, the first being to identify relevant needs and constraints, and the second to identify resources and capabilities.

### Relevant Needs and Constraints

Briefly analyzing the current system and the context of use is extremely important to ensure the right product is identified. The following questions will guide the team to focus their efforts:

- What is the problem that needs to be solved, and for whom?
- What is the existing level of risk in the current context of use?
- What are the perceptions of AI in the organization? What about the expected end-users' perceptions?
- What are the potential impacts of a new system – both beneficial and harmful?

Effective adoption of any type of software, including AI systems, requires basic knowledge of the existing context of use and the people who will use the system and then matching this with a system that provides value. "Business value continues to be a challenge for organizations when it comes to AI," according to Leinar Ramos, senior director analyst at Gartner (2024). The current processes should be reviewed to identify what is working well and what is not. That information forms a baseline of performance that acquisition teams can use to make informed decisions about potential systems. This effort may also lead to identifying areas that may need more consideration.

Individual perceptions of AI can be the most critical factor in how successful a new AI system will be. End-users who have extremely high expectations for the system may determine that oversight is not necessary. Overtrust of an AI system can result in the system being used for tasks it was not designed to do. Failures due to overtrust can be simply frustrating or, at their worst, can lead to disastrous situations such as described in Dastin (2018), Smiley (2022), and Armstrong (2023).

Under some conditions, AI tools may in fact limit, rather than enhance, scientific understanding (Messeri & Crockett, 2024). For example, scientists using AI tools for research may falsely believe they are exploring a space of all testable hypotheses, whereas they are actually exploring a narrower space of hypotheses that are testable with AI tools (Messeri & Crockett, 2024). Or they could become vulnerable to an illusion of objectivity, in which they falsely believe that AI tools do not have a standpoint or are able to represent all possible standpoints (Messeri & Crockett, 2024).

The addition of an AI system can supercharge an organization and significantly augment individual productivity. Along with these benefits, the dynamic nature of AI increases the level of risk that must be accepted by those using the system. An AI system can also increase risk for those affected by decisions made with or by the system. A system that adds more risk or requires additional fact checking may not be appropriate in contexts when decisions need to be made quickly or when correct outputs are required.

The organization's norms and culture are an important aspect of successful adoption. Engaging end-users and the team that will manage the system in a brief brainstorming activity to consider "What Could Go Wrong?" (Martelaro & Ju, 2020) will support the identification of risk for the system, increase understanding of the context of use, and can be a method to reduce fear by exploring difficult topics. User experience (UX) activities such as "Black Mirror" Episodes (C. Fiesler, personal communication, 2018) and Abusability Testing (D. Brown, personal communication, 2019) can also support these goals. Each of these methods entails minimal effort and will make a positive impact on the quality of the system and its adoption. As a reminder, these methods are not sufficient for high-risk system acquisition.

**Resources and Capabilities**

A brief analysis of resources such as data and staffing will also support the identification of the right system. Use the following questions to guide the team:

- How representative is the training dataset to the intended operational context?
- What experience does the organization have managing complex systems?

- What resources are available (or needed) for monitoring and managing the AI system over the expected period of operation?
- What type of AI system(s) are a good match with the problem to be solved?

An AI system is most effective when it is trained on data that fits the use case. A quick review of the data the team intends to use will be helpful in preparing for an AI system. For example, at the SEI, we quickly found that a computer vision system trained to identify tanks in a lush green location was not useful in identifying tanks in a desert by doing a relatively small experiment. Exploratory data analysis (U.S. Environmental Protection Agency, 2025) combined with qualitative methods can support data understanding. There are also sources of guidance from the DoD and others to support a data-driven culture, such as Gebru et al. (2021), Defense Information Systems Agency (DISA, 2025), and DrivenData (n.d.).

Any data that is used for training or otherwise contained within the system has the potential to be obtained through use of the system. Guardrails and other precautions are helpful and will work in most situations, but if undesirable or harmful information, personally identifiable information, or other types of non-public data are potentially in the system, the team will need to accept the risk of exposing that data. This is a particular risk with systems using generative AI which are designed to generate new combinations of data. These systems provide many opportunities and benefits and can be fine-tuned with additional information about a specific topic; however, they do not reliably retrieve specific data, nor are they likely to successfully reproduce the same outputs. Additionally, many claims made today about what generative AI can do are overhyped (Carleton et al., 2025).

Teams will be most successful integrating and managing AI systems when they have previous experience managing complex systems, strong technical capabilities, and a desire to learn. Cybersecurity is unlikely to be affected directly, but nearly all other aspects of the existing systems, applications, and integrations will likely be affected. Similarly, the teams' preparation for monitoring and managing the AI system will influence the systems successful adoption. "As organizations scale AI, they need to consider the total cost of ownership of their projects, as well as the wide spectrum of benefits beyond productivity improvement," said Ramos (Gartner, 2024).

As with any software system, as previously mentioned, the acquisition team needs to understand the use case and be provided with clear criteria for purchase selection. help the procurement team narrow the choices and assess products for potential suitability for its intended purpose. The new system should improve the situation and perform at least as well as the previous system.

Selecting an appropriate AI system for the problem to be solved can be a challenge, as there are many types of systems and each has strengths and weaknesses. For example, generative AI systems such as large language models (LLMs) are very popular currently, but they are not the right choice for every situation (Tao, 2024). An LLM can be an excellent solution to meet the needs for a chatbot or generating content, but it is not a good solution for decision intelligence or forecasting which require tools that can retrieve specific information or analyze data. A well-defined problem to solve will enable easier matching to an AI solution.

## Operational Experimentation

Once the initial vendors and AI system selections are identified, it is time to validate suitability, and the only way to get UX design right is to test it (Moran, 2024). Operational experiments can be conducted with a full AI system, a minimum viable product (McDonald, 2023), or a clickable (low code) prototype. The AI system does not need to be fully functional or fully integrated into the environment, but it does need to at least provide an understandable and

representative experience of the primary tasks it is intended for. This can be challenging for vendors but is a reasonable request for a substantial investment.

The people who will use the product (end-users such as warfighters, analysts, operators, etc.) need to be given access to the system or prototype in their typical environment. With just a short introduction to the system, the end-users should be able to use the system to do the core tasks it is expected to support. They should be able to interpret the output of the AI product and be able to determine if the system is working as expected. This activity is akin to usability testing (Moran, 2024), and if the system is well designed, the end-users should need only minimal support and not specific direction.

This part of the assessment is subjective by design to enable identification of failures that will erode trustworthiness as early as possible. End-users are typically the ones to discover AI technology failures, and those negative experiences are risk indicators of deteriorating trustworthiness (Gardner et al., 2023). Organizations employing these systems must therefore ensure that end-users are supported with:

- indicators within the system when it is not functioning as expected
- ability to report when the system is deteriorating or not operating properly
- information to align their expectations and needs with the potential risk the system introduces (Gardner et al., 2023)

Before determining whether to employ a new AI technology, ask these questions (Gardner et al., 2023):

- What are the limitations of the system's functionality?
- What are the safety controls to prevent this system from causing damage? How can these controls be tested?
- Is the development team able to understand and audit the output of the tool?
- How was the model trained? Could an expert retrain this tool to meet changing needs (e.g., to adhere to a new policy or to integrate new information)?
- Does the vendor enable operational experimentation and iterative phases of work?

If the operational experimentation is successful, with the success criteria met, and end-users deeming the system to be effective, then procurement can choose whether or not to consider other systems. The operational experimentation may be unsuccessful for a variety of reasons, such as the end-users not being able to complete their tasks on their own, the system not providing them with confidence that it was able to support their needs, or the system seemed untrustworthy. In these cases, the acquisition team should eliminate the system from consideration and move on to the next solution. This process provides quick and relevant feedback to the procurement team and reduces the chances of wasting funds on the wrong AI system.

## Continuity of Capabilities

Implementation of a new AI system is just the beginning. The capabilities must continue to be available to the workforce for the system to be successfully adopted and integrated into existing processes. Systems are typically rejected when the interface design and interactions diverge from the end-users needs. Connecting with end-users ensures that aspects of design such as trustworthiness and transparency are interpreted and implemented appropriately. When end-users understand the systems' capabilities and limitations and are confident using it in context, it is likely to be successful.

The workforce will need to define processes and responsibilities for the following aspects of the system (Gardner et al., 2023):

- Continuous monitoring, test, evaluation, verification and validation practices such as Derr et al. (2025) and NIST (2023)
- Continuous performance monitoring appropriate for developers and end-users
- Risk mitigation planning and implementation
- Operations for training, fine-tuning, auditing, etc. of data and models such as DeCapria (2024)

As the AI system is adopted, the workforce may need to develop guidance for productive system use and even specify the systems' recommended uses and limitations. The user and development needs will change over time, so keeping an agile mindset will help the team to respond as needed.

## Conclusion

This streamlined practice to inform procurement teams, enable operational experimentation, and ensure the continuity of capabilities will enable the Navy to more quickly implement effective AI systems that connect and augment human abilities. Better AI system procurement practices will enable the United States to solidify its position as the leader in AI and secure a brighter future for all Americans.

## References


Armstrong, K. (2023, May 27). *ChatGPT: US lawyer admits using AI for case research.* BBC. https://www.bbc.com/news/world-us-canada-65735769

Carleton, A., Ivers, J., Ozkaya, I., Robert, J., Schmidt, D., & Zhang, S. (2025, February 27). *Perspectives on generative AI in software engineering and acquisition*. https://doi.org/10.58012/gtk0-b084

Dastin, J. (2018, October 10). *Insight - Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/

DeCapria, D. (2024, November 4). *Introduction to MLOps: Bridging machine learning and operations*. https://doi.org/10.58012/zkgg-en08

Defense Acquisition University. (2020). *Software acquisition*. DAU Adaptive Acquisition Framework https://aaf.dau.edu/aaf/software/

Defense Information Systems Agency. (2025). *DISA data strategy FY25-27*. https:///DISA-FY25-27-Data-Strategy

Derr, A., Echeverría, S., Maffey , K. R., & Lewis, G. (2025, February 17). *Introducing MLTE: A systems approach to machine learning test and evaluation*. https://doi.org/10.58012/cqzv-wj37

DrivenData. (n.d.). *Deon ethics checklist*. Retrieved June 20, 2024, from https://deon.drivendata.org/

Gardner, C., Robinson, K.-M., Smith, C. J., & Steiner, A. (2023, July 17). *Contextualizing end-user needs: How to measure the trustworthiness of an AI system*. Carnegie Mellon University, Software Engineering Institute. https://doi.org/10.58012/8b0v-mq84

Gartner. (2024). *Gartner survey finds generative AI is now the most frequently deployed*. https://www.gartner.com/en/newsroom/press-releases/2024-05-07-gartner-survey-finds-generative-ai-is-now-the-most-frequently-deployed-ai-solution-in-organizations

Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., & Crawford, K. (2021, December 1). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. doi:10.1145/3458723

Hegseth, P. (2025, March 10). *Modern software acquisition to speed delivery, boost warfighter lethality.* DoD. https://www.defense.gov/News/News-Stories/Article/Article/4114775/modern-software-acquisition-to-speed-delivery-boost-warfighter-lethality/

Martelaro, N., & Ju, W. (2020). What could go wrong? Exploring the downsides of autonomous vehicles. *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '20).* Association for Computing Machinery. doi:https://doi.org/10.1145/3409251.3411734

McDonald, K. (2023). *What is a minimum viable product (MVP)?* Agile Alliance. https://www.agilealliance.org/glossary/mvp/

Messeri, L., & Crockett, M. (2024, March 6). Artificial intelligence and illusions of understanding in scientific research. *Nature, 627*, 49–58. https://doi.org/10.1038/s41586-024-07146-0

Moran, K. (2024). *Usability (user) testing 101*. Nielsen Norman Group. https://www.nngroup.com/articles/usability-testing-101/

NIST. (2023). *NIST AI RMF: Artificial intelligence risk management framework (AI RMF 1.0).* https://doi.org/10.6028/NIST.AI.100-1

Smiley, L. (2022, March 8). *'I'm the operator': The aftermath of a self-driving tragedy*. Wired. https://www.wired.com/story/uber-self-driving-car-fatal-crash/

Tao, C. (2024, August). *Do not use LLM or generative AI for these use cases*. Towards AI. https://pub.towardsai.net/do-not-use-llm-or-generative-ai-for-these-use-cases-a819ae2d9779

U.S. Environmental Protection Agency. (2025, February 13). *Exploratory data analysis*. https://www.epa.gov/caddis/exploratory-data-analysis


Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net