

# CROSSTALK

*The Journal of Defense Software Engineering*

May 2024



# AI

## PART ONE

## TAMING THE BEAST



# CrossTalk: The Journal of Defense Software Engineering

**CrossTalk: The Journal of Defense Software Engineering** is sponsored by the United States Air Force (USAF) 309 Software Engineering Group (SWEG) at Ogden Air Logistics Complex (ALC). It is also supported by other partners within the Department of Defense (DoD), other USAF systems, and the software engineering community.

The USAF 309 SWEG also publishes CrossTalk, provides editorial oversight, and technical review of the journal. The mission of CrossTalk is to encourage the engineering development and proper management of software to improve the reliability, sustainability, and responsiveness of our warfighting capability.

**CrossTalk Online:** Current and past issues are posted at the following two locations. The All Partners Access Network (APAN) and Defense Technical Information Center (DTIC)

<https://community.apan.org/wg/crosstalk/>

<https://www.dodtechipedia.mil/dodwiki/x/HwDqFQ> (Requires .mil domain for full support)

The APAN site requires registration to be a regular member and to interact with others. The DTIC website includes reports designated as unclassified and unlimited information and can be reached at: <https://discover.dtic.mil/technical-reports/>

**Subscriptions:** Please send an email to the publisher to receive a notification when each new issue is published online.

**Article Submissions:** We welcome articles of interest to the defense software community. Articles must be approved by the Technical Review Board (TRB) prior to publication. Please follow the *Author Guidelines*, available at either of the two sites above. CrossTalk does not pay for submissions. Published articles remain the property of the authors and may be submitted to other publications. Security agency releases, clearances, and public affairs office approvals are the sole responsibility of the authors and their organizations. Potential articles can be emailed to: 517SMXS.Crosstalk.Articles@us.af.mil

**Reprints:** Permission to reprint or post articles must be requested from the author or the copyright holder and coordinated with CrossTalk.

**Trademarks and Endorsements:** CrossTalk is an authorized publication for members of the DoD. Contents are not necessarily the official views of, or endorsed by, the U.S. government, the DoD, the sponsors, or co-sponsors. All product names referenced in this issue are trademarks of their respective companies.

**Publishing Schedule and Back Issues:** CrossTalk is currently being published quarterly. Please phone or email us to see if back issues are available, free of charge.

ISSN 2160-1577 (prior print versions); ISSN 2160-1593 (online)

## CrossTalk Staff

### Sponsor

James L. Diamond Jr.

### Managing Director

Christian J. Durain

### Assistant Director

Malissa Jones

### Managing Publishers

Lennis L. Burton

Siria L. Snounou

Destinie Comeau

### Technical Reviewer

Alan Sorensen

## Contact us

### Phone

Lennis L. Burton, (801) 775-3262

Siria L. Snounou, (801) 777-4734

Destinie Comeau, (801) 775-3246

### E-Mail

517SMXS.CrossTalk.Articles@us.af.mil

## SWEG Socials



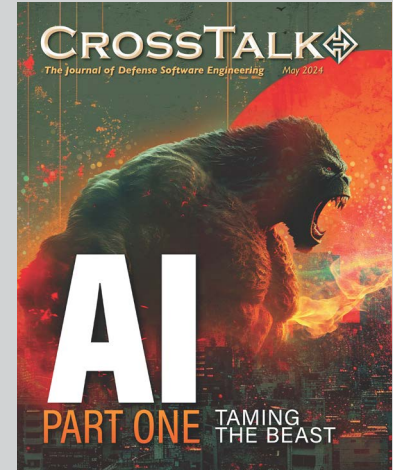
*Connect with us at  
the 309th SWEG socials*



# CrossTalk: The Journal of Defense Software Engineering

## DIRECTORY

- 4** From the Sponsor
- 8** Call for Articles
- 9** AI Part 1
- 68** Spotlight
- 72** BackTalk



Cover Design by Kent Bingham

## AI PART 1

- 9** **Assessing Risk of Using AI in DoD Weapons Systems, By Harrell J. Van Norman**  
Explains current climate of AI and assesses risks and benefits of using AI in DoD weapon systems.
- 26** **The Future of Software Engineering and Acquisition with Generative AI, By Software Engineering Institute**  
Examines the transformative potential of generative AI in redefining software engineering and acquisition practices.
- 45** **U.S Air Force Depot Operating Efficiency and Mission Readiness Using AI/ML, By Abeezer Tyebji and Vik Chauhan**  
Responds to challenges, such as extending the life of aircrafts beyond intended design, by proposing the utilization of breakthrough data science technologies.
- 53** **Imperfect Mates: Humans and AI in the Cockpit, By Maj. Richard C. Agbeyibor and Dr. Karen M. Feigh**  
Details the relationship between human and AI counterparts in military vehicles and operations.

From the Sponsor

# AI in the DoD: Power, Potential, and Risks

Eileen Wrubel

Technical Director - Transforming Software Acquisition  
Practice and Policy Directorate,  
Carnegie Mellon University

[A]



I am a product of Generation X, and my first real memorable introductions to the idea of Artificial Intelligence (AI) came from the legends of two particular blockbuster films: 1983's *War Games* and 1984's *The Terminator*.

Both films told stories of Artificial Intelligence systems run amuck, with catastrophic (or potentially catastrophic) nuclear annihilation consequences. Kids of my vintage definitely grew up with some healthy skepticism of the consequences of irresponsible development of Artificial Intelligence technologies, against the very real backdrop of Cold War nuclear proliferation. And at the same time, the wild futures imagined for us in science fiction inspired a generation of kids who grew up learning how to program on an Atari 800 or Commodore 64, with big ideas about how we could potentially change the world with AI (right after we ate our TV dinner and watched the latest episode of *Growing Pains*, of course).

Fast forward a few decades, and the future has arrived. The explosive proliferation of Artificial Intelligence technologies in the last decade has profoundly reshaped how we interact with and how we perceive the world around us. Advances in AI enable rapid analysis of complex mountains of data to empower actions and decision-making in virtually every industry. In her keynote on “The Global AI Contest” at the Advantage DoD 2024: Defense Data & AI Symposium, Deputy Secretary of Defense Kathleen Hicks emphasized the imperative in the Department of Defense (DoD) for harnessing powerful advances in AI and machine learning technology in developing and deploying mission capabilities to maintain our strategic advantage on the world stage, and also to transform and improve the operation of the DoD itself.

*“Of course, increasingly over the last dozen years or so, advances in machine learning have brought new generations of AI innovation. Much of it, happening outside of government.*

*So our task in the Defense Department today is to adopt those innovations wherever they can add the most military value. And that’s been a priority for Secretary Austin and for me since day one.*

*We knew we had to build rapidly and responsibly — iterating and investing to deliver a more modernized data-driven and AI-empowered military now.*

*There was no debating the “why” — it’s because these technologies give us an even better decision advantage than we already have today. And that’s imperative given the pacing challenge we face from the People’s Republic of China.*

*In deterring and defending against aggression, AI-enabled systems can greatly improve the speed, quality, and accuracy of commanders’ decisions — which can be decisive in deterring a fight, and winning a fight.*

*But also in managing the world’s largest enterprise. We’ve got nearly 3 million people on the payroll; a health care system serving over 9 million troops, retirees, and family members; and assets spread out worldwide over 25 million acres, roughly the size of Kentucky. The value of DoD’s assets worldwide is larger than Amazon, Apple, Microsoft, and Walmart combined — by a long shot. At this scale, we must leverage data and AI to be smarter, faster, and better stewards of taxpayer dollars” [1].*

Exploiting the power and potential of Artificial Intelligence in service of these imperatives opens up a seeming infinity of exciting questions for program managers, researchers, engineers, practitioners, and leaders at all levels of the DoD. Many (most) of us are not experts in the fields of AI and machine learning, and we are coming to terms with the vocabulary of these technologies right alongside their vast potential – and risks. Here’s an illustrative (and in no way exhaustive) sampling of the kinds of questions we face as we seek to harness these rapidly evolving technologies:

When we explore the feasibility of using AI tools and AI-enabled systems to address mission challenges – including those associated with conducting the business of managing the DoD enterprise itself - we consider questions such as:

- Is this problem a good candidate for addressing with AI?
- What AI models and techniques are appropriate to solving a particular mission challenge?
- What intellectual property rights do we need for both data and AI models?
- How do we know we can trust the results?
- How do we monitor performance in operations?

When we consider the engineering of software and AI-enabled systems, researchers in the DoD, Federally Funded Research and Development Centers (FFRDCs), academia, and industry are exploring questions such as:

- How do we need to adapt traditional software engineering and test methods to build and test AI components and AI-enabled systems?
- How can we use new AI technologies to develop new tools and techniques for building and testing software and systems?
- How can we use AI technologies to improve our cybersecurity posture or mitigate cyber threats?

We also consider new risks, threat vectors, and potential for harm associated with them:

- Do we understand the source/provenance of training data, and how training and operational data could be compromised?
- What is the potential harm from such compromises?
- How do we identify and respond to cyber vulnerabilities in AI tools and AI-enabled systems?

In this issue of CrossTalk, we hear from professionals across the defense community about addressing these kinds of questions and more as we seek to harness AI for national strategic advantage:

Harrell J. Van Norman explains the current climate of AI and assesses risks and benefits of using AI in DoD weapon systems.

John E. Robert, James Ivers, Doug Schmidt, Ipek Ozkaya, and Shen Zhang examine the transformative potential of generative AI in redefining software engineering and acquisition practices.

Vik Chauhan and Abeezar Tyebji respond to challenges, such as extending the life of aircrafts beyond intended design, by proposing the utilization of breakthrough data science technologies (AI/ML).

Richard C. Agbeyibor and Karen M. Feigh detail the relationship between human and AI counterparts in military vehicles and operations.

Unlocking the potential for AI to benefit DoD missions, operations, and people is an exciting and multi-faceted goal. I'm hopeful you'll enjoy the insights offered by our authors in this issue, and that you're eager to come back soon: The next issue of CrossTalk will feature more perspectives on the potential, challenges, and risks of AI technologies.

**- Eileen Wrubel, Technical Director, Carnegie Mellon University**

## Picture References

[A] "Terminator - Endoskeleton - Desktop Nexus Wallpapers." Desktop Nexus Wallpapers, 6 Mar. 2018, [abstract.desktopnexus.com/wallpaper/2357002](https://abstract.desktopnexus.com/wallpaper/2357002).

## References

[1] U.S. Department of Defense. "Remarks by Deputy Secretary of Defense Kathleen Hicks Keynote on 'The.'" U.S. Department of Defense, [www.defense.gov/News/Speeches/Speech/Article/3683202/remarks-by-deputy-secretary-of-defense-kathleen-hicks-keynote-on-the-global-ai](https://www.defense.gov/News/Speeches/Speech/Article/3683202/remarks-by-deputy-secretary-of-defense-kathleen-hicks-keynote-on-the-global-ai).

---

Contributions to this issue of CrossTalk by staff of the Software Engineering Institute, a federally funded research and development center, were made with funding and support from the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University (DM24-0608).

# From the Cover Designer:



There are universal themes throughout the *Godzilla vs. King Kong* movie “MonsterVerse” that are great metaphors for the emergence of AI and all the issues it brings with it [1].

A major theme of the “MonsterVerse” is how humanity needs to learn to coexist with nature and attempt to work with it rather than work against it. That same concept could also apply to humans and Artificial Intelligence. AI is expanding exponentially at a breakneck speed, and there is a growing fear that its potential power is destructive and will destroy jobs, industries, ignore

copyright and privacy, and wreak uncontrolled havoc on society.

In the movies, Godzilla actually has a role in protecting the Earth and they show that mankind needs to view him as an ally and work with him in order to defend the world from being destroyed. Since AI is here to stay, we need to be able to learn to coexist with it, but to also ensure we have the necessary safeguards to ensure it's used wisely.

**- Kent Bingham,**

**Visual Information Specialist, Hill Air Force Base**

## References

[1] *Godzilla vs. Kong*. Directed by Adam Wingard, Warner Bros., 2021.

# Call For Articles

If your experience or research has produced information that could be useful to others, Crosstalk can get the word out. We are specifically looking for articles on software-related topics to supplement upcoming theme issues. Below is the submittal schedule for the areas of emphasis we are looking for.

## AI PART 2

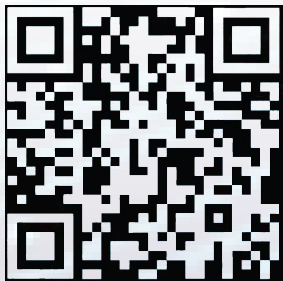
August 2024 Issue

Submission Deadline:  
June 15, 2024

## BIG DATA

Nov 2024 Issue

Submission Deadline:  
September 15, 2023



## KEEPING UP WITH THE CLOUD

February 2025

Submission Deadline:  
October 31, 2024

Please follow the Author Guidelines for Crosstalk, available at the APAN or DTIC site.

We accept article submissions on software-related topics at any time, along with Letters to the Editor, Open Forum, and BackTalk. To learn more about the types of articles we're looking for, please visit the above sites or contact us by email or phone

## Contact Us

By phone

By email

Lennis L. Burton, (801) 775-3262  
Siria L. Snounou, (801) 777-4734  
Destinie Comeau, (801) 775-3246

517SMXS.CrossTalk.Articles@us.af.mil





# Assessing Risk of Using AI in DoD Weapon Systems

HARRELL J. VAN NORMAN  
CYBER SURVIVABILITY TECHNICAL EXPERT/SCA,  
USAF AIRCRAFT SYSTEMS

## Executive Summary

The future will be dominated by Artificial Intelligence (AI); however, few strides have been made to understand the risk of its responsible use in Department of Defense (DoD) weapon systems. This paper seeks to address this deficiency by providing an approach to identify and quantify the risks associated with this ongoing phenomenon that the DoD is eager to adopt and embrace.

AI can bring many breakthroughs for DoD capabilities because of the increased efficiency, improved information gathering for decision making, and the reduction of human error and benefits in automated and autonomous systems. This is true for all foreign militaries as well, which is why it needs to be an integral part of the DoD's strategy going forward. However, risk identification, quantification, and mitigation must be conducted to effectively apply AI technologies.

At the intersection of the likelihood of a threat agent exploiting vulnerabilities with system and mission impacts, risk can graphically be depicted using the DoD standard 5x5 risk assessment matrix. The National Institute of Science & Technology (NIST) AI Risk Management Framework (RMF) helps design, develop and evaluate AI products by framing risk and trustworthiness based on four core functions: Govern, Map, Measure, and Manage. From these functions, 19 categories and 72 sub-categories provide a thorough guide to properly identify AI risk. To tailor the framework for the DoD's needs, scoring can be done using tiers as espoused in the NIST Cybersecurity Framework (CSF) to provide a rating on the likelihood of the failure to occur in each of the subcategories. With proper scrutiny, this new approach to AI risk assessment seeks to maximize the benefits of AI for the warfighter.

## The Push for AI

The DoD is acquiring, developing, and adopting AI technology throughout its agencies. In November 2023, department-wide updated guidance, Data Analytics and AI Adoption Strategy, directs DoD components to accelerate adopting AI capabilities to ensure warfighting superiority now and in the future

[1]. Adopting AI ecosystems enables superior battlespace awareness and understanding; elastic force planning and sustainment; efficient operations; and fast, precise, and resilient kill chains. Taken together, these goals will support the “DoD AI Hierarchy of Needs” which the strategy defines as: quality data, governance, insightful analytics and metrics, assurance, and responsible AI. However, few studies have attempted to adequately identify AI risk in DoD weapon systems and provide an effective approach for risk quantification. This paper aims to fill that void.

Generative AI, employing algorithms and models that generate new data, has emerged as the most rapidly adopted technology in history, faster than smartphones, tablets, or the internet [2]. The DoD’s AI strategy is to apply this emerging technology to solidify the United States’ competitive advantage across the military spectrum in over 180 instances where generative AI could add value, such as optimizing logistics, predictive maintenance, assessing battle damage, processing intelligence data, threat recognition, and automating software development. The future will be dominated by AI. Dr. Craig Martell, who formerly led the department’s Chief Digital and Artificial Intelligence Office (CDAO), advocates,

***“Rather than identify a handful of AI-enabled warfighting capabilities that will beat our adversaries, our strategy outlines the approach to strengthening the organizational environment within which people can continuously deploy data, analytics and AI capabilities for enduring decision advantage.”***

The AI frontier, characterized as the next industrial revolution, can provide decisive advantages to deter adversaries and win in a fight. In addition to aggressive adoption, we must be committed to safety and responsibility. This paper presents an effective approach to identify and quantify risk associated with the AI components integrated into weapon systems.

## What It Is

What is AI and how does it work? As the name implies, AI systems apply computational tools to address tasks traditionally requiring human analysis. This includes recognizing patterns, creating predictions, making decisions, and/or generating new content without being explicitly programmed to do so. AI is typically created with machine learning (ML) methods like supervised learning (predicts using labeled training datasets), unsupervised learning (builds associations using unlabeled datasets), and reinforcement learning (develops actions using sensors based on policies and goals). AI builds complex neural networks that are typically inscrutable to achieve these deep learning approaches. Data is at the core of all AI, but all systems have some form of bias. Bias can be unintentional or have purpose and be helpful. A goal of responsible AI is to reduce unintended, unwanted, and/or harmful bias.

Before plunging into a deeper discussion of risk assessment approaches for AI capabilities, answering the question, “How does AI differ from simple automation?” will help frame the problem. Automation uses technology to perform tasks with minimal human intervention. This includes simple, repetitive tasks like manufacturing assembly line production, or more complex processes like scheduling and data entry in an office setting. AI goes a step further by enabling machines to learn from data, adapt to new inputs, and perform human-like tasks. AI can make deterministic decisions, but its strength is making probabilistic calculations. AI examples include speech recognition, language translation, autonomous vehicle operations, predictive maintenance, intelligence gathering, and threat

recognition. AI systems make decisions and take actions based upon an understanding of the environment. In summary, automation focuses on streamlining processes, while AI involves the ability of machines to perform tasks that are analogous to those performed by humans, including the ability to think critically, make decisions, and increase productivity.

Conceptually, there are three types, levels, or stages of AI that range from weak to strong to super AI; however, today, all fielded AI capabilities are categorized as weak or narrow. Weak AI only performs a narrowly defined set of specific tasks (like Siri, Alexa, or ChatGPT) since they're limited to a single task. Facial recognition, internet searches, and self-driving cars are other examples of narrow or weak AI. Theoretically, in the future, the concept of Strong AI will provide capabilities to understand the world as a human would and learn to perform any intellectual task that a human could. And even further into the theoretical realm are Superintelligent AI systems with futuristic intelligence that can exceed the cognitive performance of humans in virtually all domains of interest. All these types of AI systems employ a diverse set of technologies, including Machine Learning, Natural Language Processing (NLP), Neural Networks (NN), Deep Learning (DL), Robotic Process Automation (RPA), Cognitive Automation (CA), and Generative AI (GAI).

## Assessing AI in DoD Weapon Systems

AI is being rapidly deployed throughout every sector of society including health care, banking, retail, and manufacturing to improve user experiences, provide enhanced services, assist in making better decisions, reduce human error, and net gains in productivity and speed. System and mission impacts within many domains that embrace AI aren't a matter of life or death. In those arenas, AI risks aren't nearly as critical as in safety-oriented domains, like healthcare and DoD Weapon Systems.

Rigorous hardware and software verification and validation through realistic developmental and operational test and evaluation is required to minimize the probability and consequences of failure. Reliability and performance risk thresholds for autonomous or semi-autonomous operations, target selection, and engagement logic are clarified in the DoDD 3000.09, Autonomy in Weapon Systems [3].

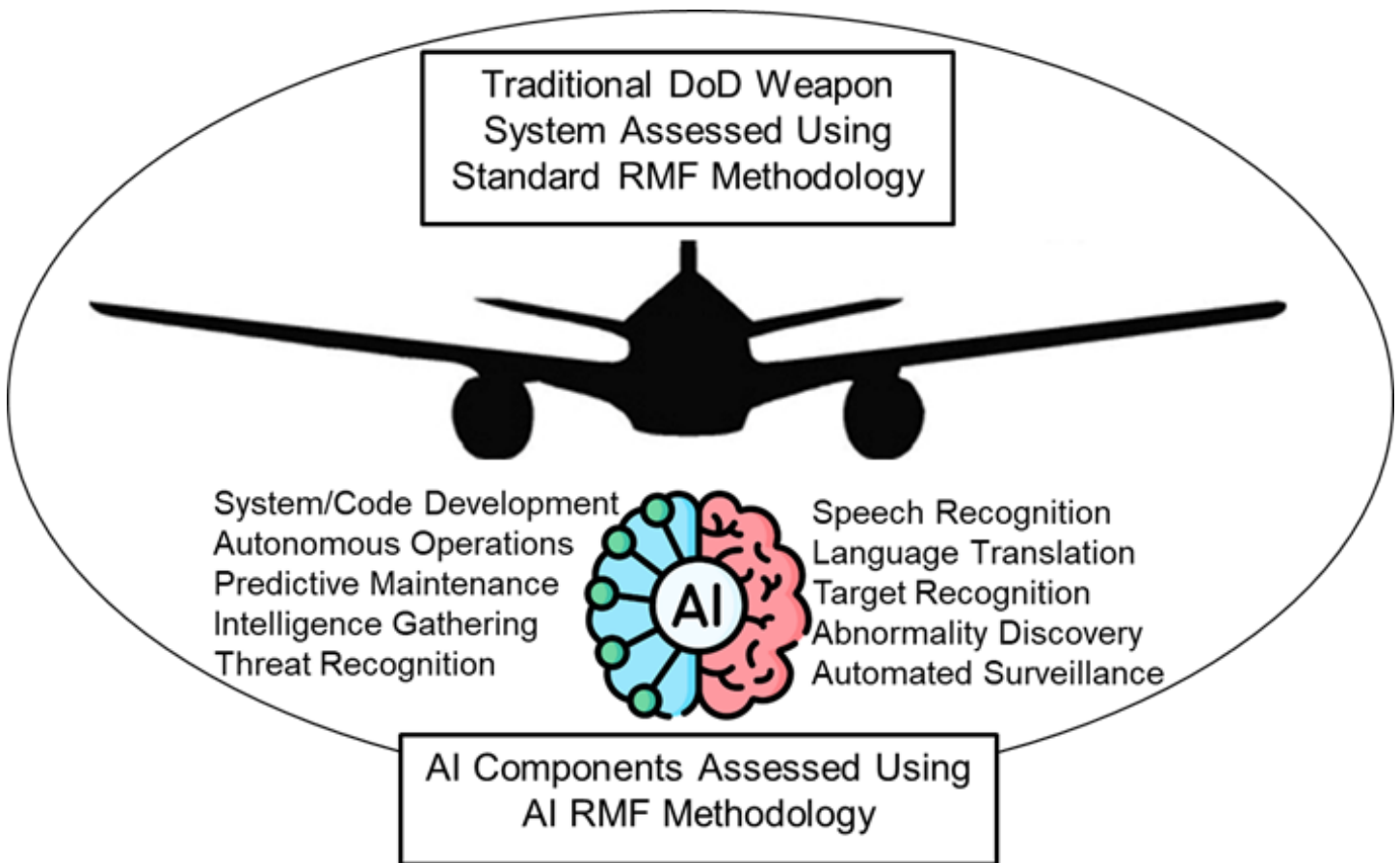
Autonomy does not dictate AI as rule-based, deterministic, expert-coded systems that have successfully supported weapon systems from their inception. Additionally, these quantifiable, predictable capabilities are much easier to certify for safety-critical systems. But AI can build upon deterministic capabilities by helping to calibrate and train AI systems. As recently demonstrated by the AI enabled Defense Advanced Research Projects Agency (DARPA) X-62A Air Combat Evolution (ACE) aircraft deploying cutting edge ML-based autonomy, AI was able to outperform manned F-16 aircraft in dynamic air-to-air combat maneuvers and missions while complying with safety and ethical norms. The X-62A ACE demonstrated that AI enabled weapon systems can provide superior performance and adaptable advantages as AI learns, improves, and adapts in situationally dynamic environments.

Automating standard business processes can provide a quick and efficient gain. But integrating AI enabled real-time autonomy into DoD Weapon Systems demands significantly greater levels of scrutiny than simply automating business processes. Nevertheless, AI will change the future of the fight, not just by making decisions quicker but also by considering alternatives and options that humans typically wouldn't. As our adversaries rapidly deploy AI capabilities, could our demise be based on over reluctance to embrace these transformative solutions?

Risk management is crucial for ensuring AI capabilities within DoD Weapon Systems are resilient in real-world settings against adversarial and non-adversarial attacks. Authorizing Officials (AOs) issuing authorization decisions based on a quantified risk posture must be neither risk adverse nor blind to risk. Scoring risk is based on the intersection of the likelihood of a threat agent exploiting

vulnerabilities with system and mission impacts. Scoring risk for AI capabilities in DoD weapon systems accounts for responsible, equitable, traceable, reliable, and governable ethical operations. First, system weaknesses must be enumerated, then risks are identified by evaluating those weaknesses to understand the system and mission impacts of potentially exploitable vulnerabilities. Next, the likelihood of threats to exploit those vulnerabilities is quantified. Finally, risk is scored based on the intersection of the likelihood of the threat exploiting the impacts of the vulnerabilities. This paper provides a basis for identifying and quantifying risk of weapon system AI capabilities.

Throughout the acquisition lifecycle, risk is managed by identifying, quantifying, and mitigating risk based on informed risk assessments. Appropriate care should be exercised to ensure potential risks are considered from the outset of the systems development lifecycle with efforts taken to mitigate potential risks and reduce the likelihood of unintended consequences throughout design, implementation, operation, and sustainment. For all components of a system that receive, process, store, or transmit information, programs should apply the RMF for Information Systems and Organizations as documented in NIST Special Publication (SP) 800-37 Revision 2 [4]. In addition, the NIST AI RMF Framework and the NIST AI RMF Playbook processes should be applied to all components of the system that integrate AI capabilities [5][6]. **Figure 1** illustrates how both traditional DoD Weapon System IT components and subsystems will be assessed by applying the standard RMF methodology and how the AI components within the DoD Weapon System should be assessed using the AI RMF methodology.



**Figure 1.** Applying both NIST SP 800-37R2 and NIST AI 1.0 RMF methodologies.

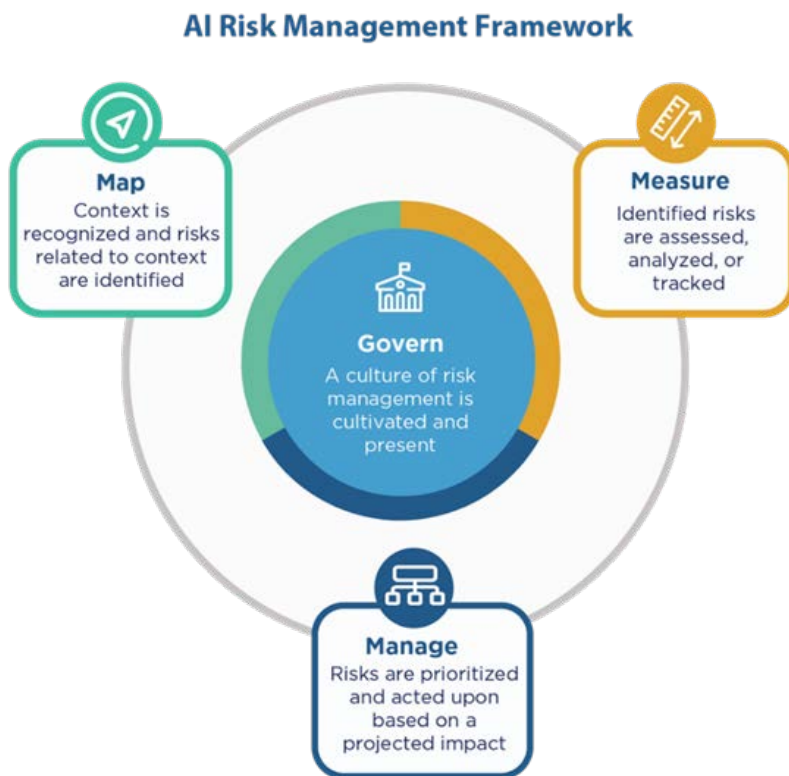
Applying multiple frameworks to assess a single system isn't an unprecedented concept as multiple frameworks are frequently applied to adequately assess risk associated with a single system. For example, the NIST Secure Software Development Framework (SSDF) is frequently applied to the developmental portion of a system while the NIST RMF is applied to assess the system in the context of testing, connectivity, and operation for Interim Authorizations To Test (IATTs), Authorizations To

Connect (ATCs), and Authorizations To Operate (ATOs). Similarly, the NIST Privacy Framework can be applied to identify and manage privacy risk when building innovative products and services while protecting individuals' privacy [7][8]. The NIST Workforce Framework for Cybersecurity is a standard approach and common language to describe cybersecurity workforce roles and responsibilities [9]. All these various frameworks provide valuable tools within the scope of their context. So, for DoD Weapon Systems with AI capabilities, they should apply NIST RMF to manage risk for the platform as a whole and apply the AI RMF to manage risks associated with the AI components and capabilities.

This unique approach for identifying and quantifying AI risk was born from the unique challenges associated with adequately assessing these capabilities. AI functions uniquely apply complex neural networks with billions of interconnected nodes across a multitude of observable input, output, and hidden layers. Trustworthiness, normalized biases, validity of training data, and adequacy of labels are uniquely difficult to assess. The dynamic nature of adaptive AI capabilities that are constantly changing, learning, and adjusting introduce even more challenges.

AI technologies are transforming society; however, these technologies also pose risks that can negatively impact individuals, groups, organizations, communities, society, the environment, and the planet. AI brings risks that are not addressed by current risk frameworks and approaches. While risk management processes generally address undesirable consequences, the AI RMF Framework offers approaches to minimize anticipated negative impacts of AI systems and identify opportunities to maximize positive benefits. Effectively managing AI risk can lead to trustworthy AI systems unleashing potential benefits to individuals, communities, and our society, but evaluating trustworthiness, bias, and risk cannot be overlooked.

## A Framework for Managing AI Risk



**Figure 2.** NIST AI Risk Management Framework four core functions.

There are several AI Frameworks; however, the NIST AI RMF Framework, with its four core functions of Govern, Map, Measure, and Manage, was adopted as the basis of this approach to identify and quantify AI risk to DoD Weapon Systems. The NIST AI RMF Framework assesses if the system is valid, reliable, secure, and resilient, including evaluations for accuracy and robustness. Explainable and interpretable models that are fair with harmful bias managed are objectives of the NIST AI RMF Framework. **Figure 2** illustrates the four core functions of the AI RMF with concise definitions describing the purpose and intent of each core function. The four core functions of Govern, Map, Measure, and Manage are decomposed into 19 categories and 72 subcategories. Governance is designed to be a cross-cutting function to inform and be infused throughout the other three functions. Mapping provides context and visibility by categorizing all the varying levels, components, capabilities, targeted

Function	Category	Subcategory
<b>Govern</b>	6	19
<b>Map</b>	5	18
<b>Measure</b>	4	22
<b>Manage</b>	4	13
<b>Total</b>	<b>19</b>	<b>72</b>

**Table 1.** NIST AI RMF core functions, categories, and subcategories.

usage, goals, expected benefits, impacts, and costs of the AI system. The Measure function provides quantitative and qualitative methods and metrics to analyze, assess, benchmark, and track AI risk and trustworthiness over time with feedback. Management is how AI risks are prioritized, responded to, monitored, communicated, documented, and improved over time. **Table 1** associates the four NIST AI RMF Core Functions with their corresponding numbers of Categories and Subcategories.

Each of the four core functions are defined below with their specific categories identified and described.

The AI RMF Govern function creates a culture of risk management to design, develop, deploy, oversee, operate, and monitor AI systems. The Govern function has six categories which translate to the following six risk candidates:

1. GOV1: Policies, processes, and procedures are in place
2. GOV2: People are empowered, responsible, and trained
3. GOV3: Workforce and processes are prioritized
4. GOV4: Communication channels are established
5. GOV5: Relevant AI actors are engaged
6. GOV6: Supply chain for third-party software and data is enforced

The AI RMF Map function establishes a context to frame risks. The Map function has the following five categories that translate to five risk candidates:

1. MAP1: Purpose, users, expectations, impacts, assumptions, limitations, actors, mission, goals, and risk tolerances are established and understood
2. MAP2: System categorized with TEVV (Test, Evaluation, Verification, & Validation) considerations identified and documented
3. MAP3: Benchmarks established for expected costs and benefits, operator proficiency, and human oversight

4. MAP4: Risks and benefits mapped for all components, including third-party software and data
5. MAP5: Likelihood and magnitude of beneficial and harmful impacts are identified

The AI RMF Measure function employs quantitative and qualitative techniques and methodologies to analyze, assess, benchmark, monitor, and document AI risk and related impacts. The Measure function has four categories that translate to the following four risk candidates:

1. MEA1: Implement an approach and metrics to measure AI risks
2. MEA2: Evaluate for trustworthiness by applying TEVV criteria for: fairness, bias, explainability, interpretability, privacy, security, resiliency, safety, transparency, accountability, validity, reliability, human factors, effectiveness, and impact
3. MEA3: Tracking continual improvement, unanticipated and emergent AI risks
4. MEA4: Efficacy of measurement and consistency

The AI RMF Manage function responds to, recovers from, and communicates about prioritized incidents or events. The Manage function has four categories that translate to the following four risk candidates:

1. MAN1: Risks are prioritized, responded to, and managed
2. MAN2: Strategies to maximize AI benefits and minimize negative impacts are planned, prepared, implemented, documented, and informed by input from relevant AI actors
3. MAN3: Monitor risks and benefits from third-party entities
4. MAN4: Risk treatments are documented and monitored

**Table 2** depicts the 19 risks categories with brief descriptions of their purpose and intent. These 19 risk categories will effectively ensure the AI system is appropriately governed, mapped, measured, and managed.

**Table 2.** 19 categories for scoring AI risk.

<b>Govern:6</b>	<b>Map:5</b>	<b>Measure:4</b>	<b>Manage:4</b>
Policies, processes, procedures, and practices are in place, transparent, and implemented effectively	Context is established and understood	Appropriate methods and metrics are identified and applied	AI risks based on assessments and other analytical output are prioritized, responded to, and managed
Accountability structures are in place so that the appropriate teams and individuals are empowered, responsible, and trained	Categorization of the AI systems are performed	AI systems are evaluated for trustworthy characteristics	Strategies to maximize AI benefits and minimize negative impacts are planned, prepared, implemented, documented, and informed by input from relevant AI actors
Workforce diversity, equity, inclusion, and accessibility processes are prioritized	AI capabilities, targeted usage, goals, and unexpected benefits and costs compared with appropriate benchmarks are understood	Mechanisms for tracking identified AI risks over time are in place	AI risks and benefits from third-party entities are managed

Organizational teams are committed to a culture that considers and communicates AI risk	Risks and benefits are mapped for all components of the AI system including third-party software and data	Feedback about efficacy of measurement is gathered and assessed	Risk treatments including response and recovery, and communication plans to the identified and measured AI risks are documented and monitored regularly
Processes are in place for robust engagement with relevant AI actors	Impacts to individuals, groups, communities, organizations, and society are characterized		
Policies and procedures are in place to address AI risks and benefits arising from third-party software, data, and other supply chain issues			

**Table 2.** 19 categories for scoring AI risk.

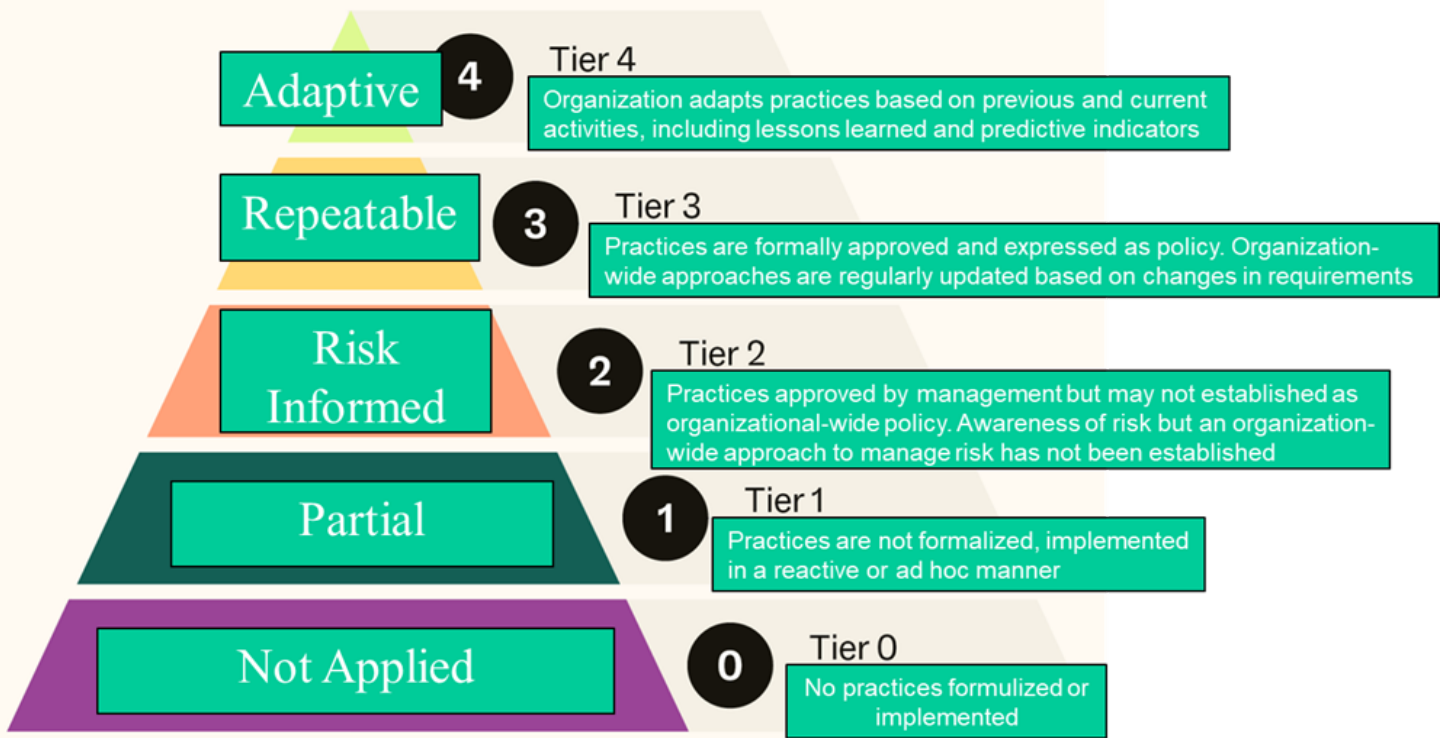
Each of these 19 categories has respective subcategory capabilities. The 72 NIST AI RMF subcategories provide useful mechanisms to identify potential risk candidates; however, it lacks a means to efficiently quantify the effectiveness of the aspects of a safe, secure, robust, and trustworthy AI system. What’s needed is a standardized means to score each of the subcategories to evaluate risk. By combining the construct of Tiers within the NIST CSF we can both identify and quantify the risk of AI Systems [10].

## Scoring Capabilities & Quantifying Risk

NIST CSF Tiers characterize the rigor of an organization’s cybersecurity risk governance and management practices, and they provide context for how an organization views cybersecurity risks and the processes in place to manage those risks. The Tiers used to rank subcategory capabilities, as shown below and in **Figure 3**, guide how to rank an organization’s currently deployed practices for managing cybersecurity risk:

- **Not Applied (Tier 0)** – There are no practices formalized or implemented in this area
- **Partial (Tier 1)** – Practices are not formalized and are implemented in a reactive or ad hoc manner
- **Risk Informed (Tier 2)** – Practices are approved by management but may not be established as organizational-wide policy. There is awareness of risk but an organization-wide approach to manage risk has not been established
- **Repeatable (Tier 3)** – Practices are formally approved and expressed as policy. Organization-wide approaches are regularly updated based on changes in requirements
- **Adaptive (Tier 4)** – The organization adapts practices based on previous and current activities, including lessons learned and predictive indicators





**Figure 3.** Ranking AI RMF subcategories via NIST CSF tiers.

The Tiers help set the overall tone for how an organization manages cybersecurity risks for this AI System and describes a progression from doing nothing to informal ad hoc responses to approaches that are risk-informed to repeatable practices that are codified in formal policy to adaptive practices that are continuously improving. Ranking each of the 72 AI Subcategories with a value from 0 to 4 will help quantify likelihood of potential compromise to the system. A chain is only as strong as its weakest link; therefore, the lowest subcategory ranking should be applied as the overall ranking for that category. This will result in 19 risk candidates corresponding to each of the 19 categories, each with a likelihood value that corresponds to its tier ranking.

<b>Likelihood</b>	L-5					
	L-4					
	L-3					
	L-2					
	L-1					
		I-1	I-2	I-3	I-4	I-5
	<b>Impact</b>					

**Figure 4.** DoD Standard 5x5 risk reporting matrix.

## Risk Reporting Example

The DoD employs a 5x5 risk assessment matrix, in accordance with DoD Directive (DoDD) 5000.1, and DoD Instruction (DoDI) 5000.2 [11][12]. This risk assessment matrix reports cyber risks based on the intersection of Likelihood of occurrence and the system/mission consequence or Impacts. The standard 5x5 DoD risk reporting matrix is illustrated in **Figure 4**. The five levels of risk correspond to the colors depicting Red as Very High, Orange as High, Yellow as Moderate, Light Green as Low, and Dark Green as Very Low.

The DoD Standard 5x5 Risk Reporting Matrix scores Likelihood of occurrence values on a scale of 1-5 based on the probability or susceptibility of compromise by a threat agent, that ranges from one end of the spectrum where L-1 corresponds to Highly Unlikely, to the opposite end where the value L-5 corresponds to Near Certain. For each of the 19 AI risk categories, use **Table 3** to map the 0-4 Tier rankings to the five Likelihood values. **Table 3** illustrates how to correlate the five Tier rankings to the five Likelihood values according to the following mapping.

Tier Ranking	Likelihood Value
0 - Not Applied	5 - Near Certain
1- Partial	4 - Highly Likely
2 - Risk Informed	3 - Somewhat Likely
3 - Repeatable	2 - Unlikely
4 - Adaptive	1 - Highly Unlikely

**Table 3.** *Correlating tier Rankings to Likelihood Values.*

Since risk is the intersection of likelihood and impact, the likelihood values need to be combined with impact level to arrive at a risk rating. The impact levels 1-1 through 1-5 depicted in **Figure 4** relate to the potential consequences of an exploit to degrade the system and/or mission according to the following scale: 1=Minimal, 2=Minor, 3=Moderate, 4=Significant or Unacceptable, 5=Severe or Catastrophic mission failure or exceptionally grave damage.

The impact level is largely influenced by the domain the system operates within. In areas where AI supports non-mission critical capabilities, the impact would be significantly lower than in deployments that do support critical capabilities and that impact safety to people’s lives and/or our nation. AI deployed in DoD Weapon Systems should be on the higher end of that spectrum. For DoD Weapon Systems, it is recommended that an impact level of 4 be applied, relating to the fact that potential degradations in capabilities would have a significant or unacceptable impact to the system and/or mission. Variations from that impact level can be made at the discretion of the system owner on a case-by-case basis.

For DoD Weapon Systems with an impact level of 4, there are three potential risk ratings: High, Moderate, and Low. This risk assessment methodology derives those risk levels from the Tier ranking of subcategories that aggregate to the 19 AI risk categories. Simply stated, if the subcategory Tier ranking of needed capabilities is 0-Not Applied or 1-Partially Applied, then the risk rating is High. When the subcategory capabilities are 2-Risk Informed, the risk rating is Moderate. When the subcategory capabilities are 3-Repeatable or 4-Adaptive, then the risk rating is Low. Whenever a risk rating is unacceptably high, then a Plan of Action & Milestones (POA&M) should be developed to effectively mitigate risk by applying the subcategory capabilities with an increased level of rigor and discipline. **Table 4** depicts risk ratings of High, Moderate, and Low from applying an impact level of 4 and the corresponding tier rankings and associated likelihood values.

Tier Ranking	Likelihood Value	Risk Ranking
0	5	High
1	4	High
2	3	Moderate
3	2	Low
4	1	Low

**Table 4.** Scoring risk based on tier ranking and likelihood values.

## Factoring Bias

For all the amazing possibilities and helpful benefits AI can provide, inherent unfair bias must be factored into the predictions, recommendations, and answers it provides. Models may exhibit good overall performance but still have unintended bias that may result in harm. The downside of AI is innate bias that misinforms and misleads decision making. The negative influence of AI bias must be mitigated by recognizing it, compensating for it, and ultimately removing it.

What is the source of bias in AI systems? Is bias born from biased programmers, biased data, or biased data labeling? Skewed training data is the predominant contributor to AI bias. Unique demographic characterizations also account for misrepresentations and misleading results.



Entirely eliminating all bias in AI systems may be theoretically impossible but it is essential to recognize unexpectedly skewed results and mitigate harmful bias as much as possible to achieve controllable and safe results. Even small divergences from expected behavior can have a “butterfly effect,” in which seemingly minor biases can be amplified and have far-reaching consequences. What practical interventions can reduce prediction bias? Correcting unfair bias requires normalizing training data and applying cross-demographic averaging.

Statistical bias can be measured as the difference between the expected value of a prediction

and the true value of the outcome variable. Representativeness bias can be measured as the distribution of the data samples across different groups using a chi-square test. Fairness bias can be measured as the disparity of a prediction for different groups or individuals.

## Measuring Trustworthiness

Trustworthiness is dependably doing well what users intend and not doing undesirable things [13]. Qualitatively and quantitatively measuring AI trustworthiness is one of the greatest challenges to properly gauge these inscrutable systems. Functional performance evaluations are useful to measure validity and reliability. NIST defines the essential components of AI trustworthiness as:

- Validity and reliability
- Safety
- Security and resiliency
- Accountability and transparency
- Explainability and interpretability
- Privacy
- Fairness with harmful bias mitigated

Building trustworthy AI systems must be considered throughout the system lifecycle to support mission-critical capabilities. Trustworthy design considerations should be embedded from the initial planning stage through release and sustainment. By intentionally building trustworthiness throughout system design, organizations can capture the full potential of AI's intended promise.

How well does the model react by recovering to the desired performance after failures? Does the model perform outside natural variations from how it was trained? Is the model proactive to effectively respond to malicious adversaries who have some knowledge and system-level access? How certain does a model's predictions need to be? Does the system need to be updated, retrained, or revalidated over time due to changes in its environment? Is the training data, testing data, and validation data representative of the operational environment? Does the operational context and associated data change over time? Are there bottlenecks that introduce latency, affecting real-time analytics or autonomous navigation? These are some of the many questions to consider as part of evaluating for trustworthiness.

Both over and under trusting an AI system can lead to regrettable outcomes, so the DoD's goal should be appropriately calibrating trust. Too much trust can endanger users who rely on it in conditions where the AI system performs poorly. Conversely, too little trust may lead users to abandon the capabilities or cause them to scrutinize every decision, driving them to become overwhelmed by doing work that was not in the Concept of Operations on top of all their other tasks and responsibilities.

## Failures in AI

The following real-life scenarios illustrate how AI deployed in DoD Weapon Systems could lead to unfortunate results. Hopefully this emphasizes the critical need for adequate risk management.

The DARPA Squad X program deployed AI to detect enemy forces in urban environments. Trained on images of soldiers walking, eight Marines were challenged to defeat the AI sensor without being detected. All succeeded by somersaulting, hiding under cardboard boxes, and camouflaging as fir trees.

A helicopter pilot is unaware that their threat recognition system does not work well in a forested environment. Their undue trust and failure to increase vigilance leads them to being shot down.

An operator does not understand that their target recognition system has a large amount of uncertainty. They mistakenly identify a school bus as an enemy troop transport.

By the time a machine learning fault-recognition system onboard an aircraft has enough data to identify the cause of a flameout, it is too late to take corrective action.

Soldiers had good experiences with their Optionally-Manned Fighting Vehicle AI driver in the lowlands and have been letting it drive unsupervised in mountainous terrain, where the AI system was untrained. It ends up driving off a cliff.

A warfighter intentionally aims to miss for a warning shot, but their AI assist doesn't understand their intent and lacks a warning shot function. It "corrects" and kills the target.

## Effective Test & Evaluation

Experienced AI testers with the right domain knowledge are a scarce and valuable commodity. This makes effective AI testing a difficult challenge. To further complicate the situation, operationally realistic testing tools are often lacking. The DoD has compiled a Responsible AI (RAI) Toolkit consisting of 70+ industry-standard, open-source tools to assist as a starting point to assess and mitigate risks or improve development of AI systems [14]. Eventually, DoD-specific tools will augment this RAI Toolkit.

An effective Test & Evaluation (T&E) strategy captures mission capabilities, prioritizes assessment areas, specifies resources required, identifies shortfalls in resources, and describes the test activities necessary to evaluate the system. The most critical ingredient for effective testing is often engagement from the warfighter.

All systems should incorporate operational realism into their T&E plans throughout the system's lifecycle. The system under test (SUT) must be production representative to accurately represent the planned fielded configuration that end users will deploy. When deployed outside of the conditions in which they were trained, AI/ML models can perform unpredictably and fail to conform to human expectations. Furthermore, when AI systems are trained with poor quality, nonrepresentative data, they will likely be ineffective. However, they may appear to be misleadingly effective if the test dataset is not operationally realistic. Testers should create adversarial examples and other robust training techniques for the test and development of AI. Correcting a complex model trained on unrepresentative data may be exceedingly challenging, if not impossible late in development. This is why operationally realistic testing early in the lifecycle is essential.

Traditional T&E methods are still applicable, but the novel challenges of non-deterministic AI systems exacerbate the capabilities of legacy testers to execute and determine when the system has been adequately tested. Testing needs to be different for AI-enabled systems; however, evaluation standards have not usually been established for tasks frequently performed by humans. Determining what constitutes adequacy for tasks without historical benchmarks makes evaluating performance challenging, especially for tasks that are not easily quantifiable. Therefore, testing must evolve to account for the challenges that AI systems impose.

What essential elements are needed for effective testing? Correctness is just the tip of the iceberg when it comes to rigorously and robustly evaluating the performance of AI models. Testing an AI model is vital for its quality, reliability, and usefulness. But ensuring testing is sufficiently robust is not simple, as many subtle aspects of performance require evaluation and validation. Correctness, the

most visible and intuitive metric, shows how well a model achieves its functional performance goals. But correctness measurements like accuracy, precision, and/or recall alone are not enough for rigorous and robust performance evaluation. At present, we face a perplexing choice between fielding AI systems of unknown trustworthiness or being bound by the limitations of our ability to provide valid and compelling evidence for trustworthiness. Developers are on the verge of having AI systems whose potential employment is limited not by their trustworthiness, but by our ability to understand and characterize that trustworthiness.

Many aspects of AI are hidden, but vital for ensuring the quality and reliability of the model. These aspects include how the model handles different sources of error (such as bias and drift), how the model explains its output and reasoning, how the model responds to different situations and inputs (such as latency and robustness), and how the model represents the real-world problem and data. Representativeness and resilience aspects are often interrelated and complex and need to be carefully considered when testing an AI model. Effective AI testing is not simple.

## Engaging Warfighters

What might seem like a sensible design choice to an engineer may not make sense to warfighters. Warfighter engagement is essential as early as possible and throughout the acquisition lifecycle to unearth unexpected anomalies. Early engagement will help shape designs to ensure operational effectiveness. Throughout the system development lifecycle, engagement is crucial to meet warfighter objectives. Post-fielding evaluations also require warfighter expertise as the system evolves.

AI systems might function as intended but are futile when they don't effectively interact with operational users. The need for deployed user interaction further justifies warfighter engagement in AI that operates, collaborates with, or coexists within DoD Weapon Systems. At the simplest level, AI capability can't be useful if the operational users ignore it or turn it off. Calibrating AI to the roles for which the capability is intended to interact is essential for effective performance.

## Summary

Continuously deploying data analytics and AI capabilities delivers an enduring decision advantage, accelerating the speed of commanders' decisions and improving the quality and accuracy of those decisions. This paper presents an innovative, standards-based approach to identify and quantify risks associated with the AI components in a DoD Weapons System. As the DoD rapidly embraces AI capabilities, it is essential that risks be managed appropriately. This novel approach will provide Authorizing Officials an effective means to identify and quantify risks so essential mitigation strategies can be employed to maximize the benefits AI provides the warfighter.

## References

[1] U.S. Department of Defense. "Data, Analytics, and Artificial Intelligence Adoption Strategy: Accelerating Decision Advantage." June 27, 2023

[2] Sidar, Cenk. "Suddenly AI: The Fastest Adopted Business Technology in History." Forbes, 5 Oct. 2023, [www.forbes.com/sites/forbestechcouncil/2023/04/05/suddenly-ai-the-fastest-adopted-business-technology-in-history/?sh=1485e74fb5c2](https://www.forbes.com/sites/forbestechcouncil/2023/04/05/suddenly-ai-the-fastest-adopted-business-technology-in-history/?sh=1485e74fb5c2).

[3] Hicks, Kathleen H. "DoD Directive 3000.09 Autonomy in Weapon Systems." DoD Directive

3000.09, 25 Jan. 2023, [www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf](http://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf).

[4] Joint Task Force. “Risk Management Framework for Information Systems and Organizations: A System Life-cycle Approach for Security and Privacy.” 1 Dec. 2018, <https://doi.org/10.6028/nist.sp.800-37r2>.

[5] Raimondo, Gina M., et al. “Artificial Intelligence Risk Management Framework (AI RMF 1.0).” NIST AI 100-1, report, Jan. 2023, [nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf](http://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf).

[6] “NIST AI RMF Playbook.” National Institute of Standards and Technology, 2023. <http://airc.nist.gov>

[7] Dodson, Donna, et al. “Secure Software Development Framework (SSDF),” Version 1.1, Recommendations for Mitigating the Risk of Software Vulnerabilities, February 2022

[8] National Institute of Standards and Technology. “NIST PRIVACY FRAMEWORK: A TOOL FOR IMPROVING PRIVACY THROUGH ENTERPRISE RISK MANAGEMENT,” VERSION 1.0. 16 Jan. 2020, [nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.01162020.pdf](http://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.01162020.pdf).

[9] Smith, Matthew C., et al. NIST Special Publication 800-181 Revision 1 “Workforce Framework for Cybersecurity (NICE Framework).” Directed by Rodney Petersen, U.S. Department of Commerce, Nov. 2020, [nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-181r1.pdf](http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-181r1.pdf).

[10] “The NIST Cybersecurity Framework (CSF) 2.0.” NIST CSWP 29, report, 26 Feb. 2024, [nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf](http://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf).

[11] Office of the Under Secretary of Defense for Acquisition and Sustainment, et al. DoD Directive 5000.01. 9 Sept. 2020, [www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/500001p.pdf](http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/500001p.pdf).

[12] Department of Defense. Operation of the Defense Acquisition System. 26 Nov. 2013, [www.acq.osd.mil/fo/docs/DSD%205000.02\\_Memo+Doc.pdf](http://www.acq.osd.mil/fo/docs/DSD%205000.02_Memo+Doc.pdf).

[13] Tate, David. “Trust, Trustworthiness, and Assurance of AI and Autonomy.” Institute for Defense Analysis, April 2021

[14] Responsible AI Toolkit. [rai.tradewindai.com/shield/intake](http://rai.tradewindai.com/shield/intake).

# About the Author



Mr. Harrell Van Norman is the cyber tech expert/SCA for USAF Aircraft Systems responsible for the assessment and authorization (RMF ATO and AW approval) of USAF Fighters, Bombers, Mobility, ISR, Executive, Agile Support Systems, and SAP Systems. He is a skilled and certified principal network design engineer with a balanced background of modeling and analysis, simulation and optimization, local and wide-area network security, and risk management. Mr. Van Norman holds a MS in Engineering from the University of Dayton and a BS in Systems Science Engineering from Michigan State University.

**Mr. Harrell Van Norman**

**Cyber Tech Expert**

**Wright Patterson AFB**

**harrell.van\_norman@us.af.mil**



HILL AIR FORCE BASE

**STEM**



**Work That Means Something**

## WHY STUDY STEM?

- Create to improve lives
- Work on a team like no other
- Give yourself thousands of opportunities— be an engineer or computer scientist
- Be an intern/earn a scholarship
- Be part of the Hill AFB Civilian STEM Workforce (*no military commitment*)

*Want to learn more or schedule a career presentation?*

*scan the QR code:*



[www.hill.af.mil/STEM](http://www.hill.af.mil/STEM)





# The 76th Software Engineering Group is hiring!

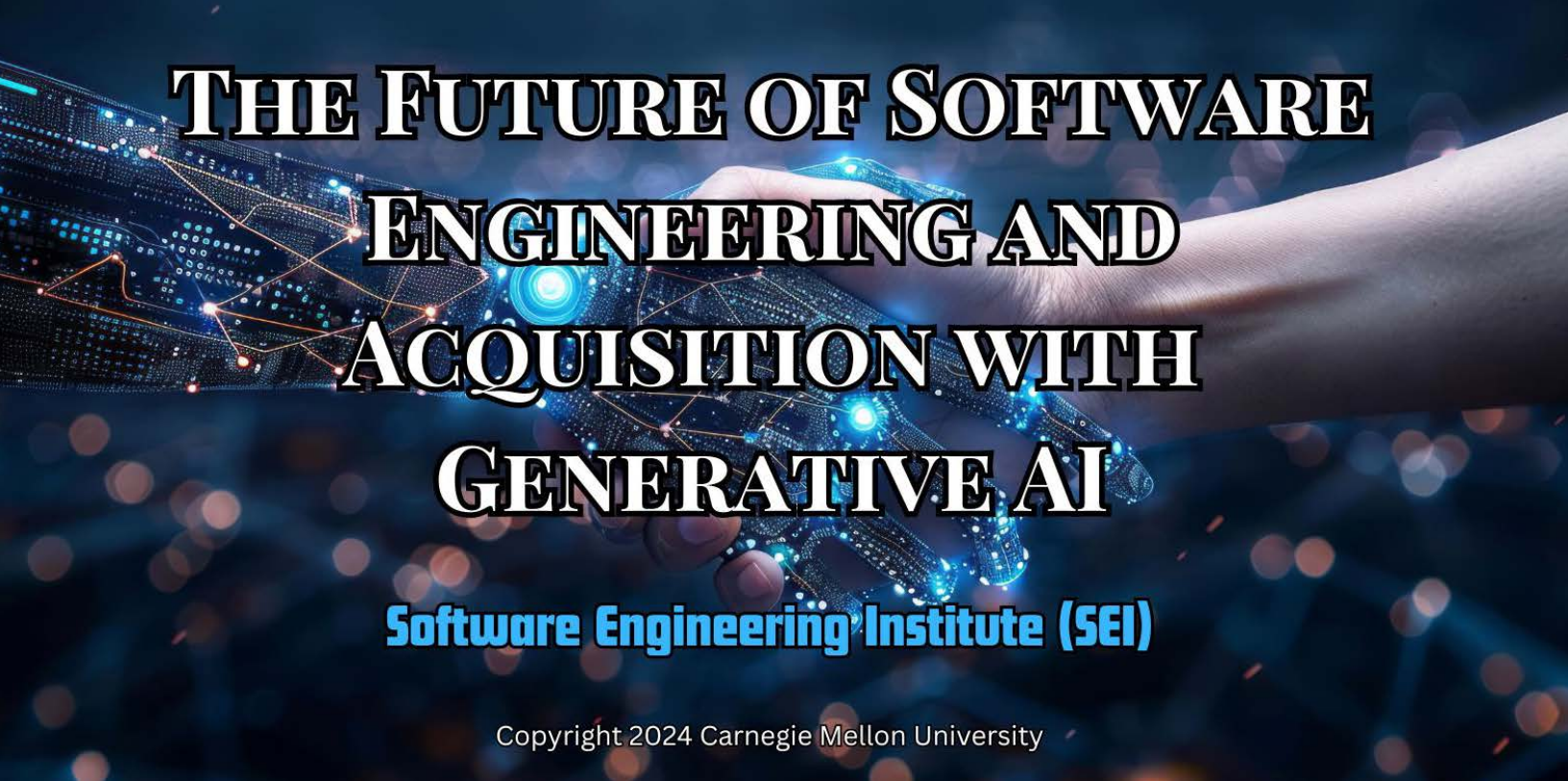
**Location:** 76 SWEG @ Tinker AFB, OK

- Electronics Engineers (0855)
- Computer Engineers (0855)
- Computer Scientists (1550)
- IT/Cybersecurity Specialists (2210)

***To apply, submit your resumes to:***  
**[AFSC.SWH.HumanCapital@us.af.mil](mailto:AFSC.SWH.HumanCapital@us.af.mil)**



**US CITIZENSHIP REQUIRED**



# THE FUTURE OF SOFTWARE ENGINEERING AND ACQUISITION WITH GENERATIVE AI

Software Engineering Institute (SEI)

Copyright 2024 Carnegie Mellon University

## Introduction

This article examines the transformative potential of generative artificial intelligence (AI) in redefining software engineering and acquisition practices. Distinguished by the ability to create new content from vast datasets, generative AI promises increased productivity and innovation that is particularly relevant for the Department of Defense (DoD). However, adopting generative AI poses both opportunities and challenges. This article delves into the nature of generative AI, focusing on large language models (LLMs) that produce text-based content and highlighting their application for tasks like code generation, document summarization and discrepancy analysis, and decision support. However, the potential pitfalls of AI, such as overfitting and biased outputs, necessitate robust validation methods and human oversight. We advocate integrating generative AI with human expertise to navigate the challenges and fully leverage its potential in software engineering and acquisition.

## Demystifying Generative AI and LLMs

Generative AI is only one form of AI, which also includes machine learning, expert systems, neural networks, fuzzy logic, evolutionary algorithms, and reinforcement learning [1]. Understanding what makes generative AI different from other forms of AI is crucial for understanding how it can best be applied to solve software engineering and acquisition problems. For example, machine learning and generative AI both rely on training sophisticated models, but these models excel at different tasks. Machine learning typically focuses on classification problems (e.g., recognizing an object within an image), whereas generative AI differs in its ability to create new content (e.g., generating answers to user questions).

A large language model is a form of generative AI that creates text-based content and has many potential applications in software engineering and acquisition, both of which are domains with extensive text-based content. At its core, an LLM is a sophisticated neural network trained on enormous repositories of data encompassing books, code, articles, and websites. Through this training, an LLM grasps the intricate patterns and interconnections within the input it's trained upon. The probabilis-

tic and randomized selection of the “next token” when generating outputs can provide users with an impression of correctness and style. Consequently, LLMs can produce coherent output, including grammatically accurate sentences and passages that closely resemble human-generated content, as well as syntactically and semantically precise software code segments.

## Challenges and Considerations

AI models are distinct from other types of models (e.g., simulation models) that encode precise mathematical or physics-based rules for a domain. AI models are statistically based, and they learn patterns from training on large data corpora. While this training allows AI models to discover relations that humans may not recognize, the statistical nature of AI models can also yield errors. Consequently, AI models face several pitfalls that include overfitting to specific datasets or failing to adapt to new and diverse data scenarios. These pitfalls underscore the need for robust validation methods to ensure these tools are enhancing, rather than compromising, the quality and reliability of software products.

LLMs, for example, are generally adept at parsing and generating nuanced text, which is valuable for generating documentation, commenting on code, and facilitating conversational interfaces within development tools [2]. The application of LLMs is not without challenges, however, since they can misrepresent context or yield biased output based on the data they were trained on. Consequently, careful human review and oversight is needed to align the text output of LLMs with software development standards, governance policies, and ethical norms.

## Opportunities

Despite these issues, incorporating generative AI—particularly LLMs—into software engineering and acquisition processes can yield a number of benefits. For example, LLMs can enhance problem-solving capabilities, streamline the creation and management of technical documentation, and foster adaptive information-centric workflows. Naturally, generative AI must be applied judiciously, with careful attention to potential biases, error margins in novel situations, the clarity of user query interpretation, and the ethical implications of their deployment.

The synergy between human expertise and generative AI in software engineering and acquisition is essential to leverage the full potential of these technologies. As generative AI continues to progress, it should not supplant human involvement but rather complement it, ensuring that AI-augmented outputs are understandable and ethically sound. It is vital to maintain human oversight to validate the reliability and accuracy of outputs produced by generative AI.

While the generative AI discussion in this paper focuses on a single modality (text) in conjunction with LLMs, applications in other modalities are maturing quickly. Generative AI can already create images, audio, and video based on text input, each of which creates additional opportunities for the application of this technology. For example, generative AI can be used to:

1. Create prompt-based prototypes [3]
2. Simulate user interface designs
3. Create educational videos that demonstrate the use of new software tools or features
4. Automate the production of training materials based on acquisition documents
5. Generate realistic audio-visual scenarios for testing software interoperability

- 6. Craft visualizations that help stakeholders understand the implications of different software architectures.

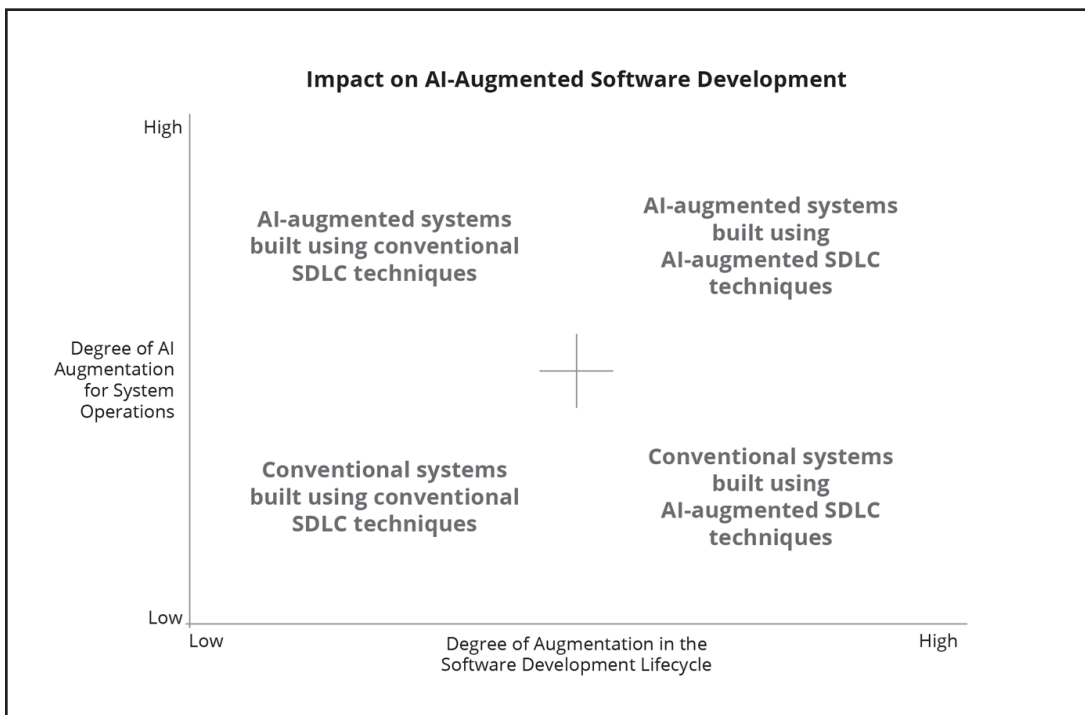
# Generative AI in Software Engineering

The integration of generative AI presents various opportunities in software engineering tasks, such as code generation, configuration deployment, and testing support, as summarized below:

- It can generate boilerplate code, significantly enhancing software development workflows by reducing manual coding errors and increasing developer productivity.
- It can generate setup and provisioning files for software environments, ensuring consistency and accuracy across configurations of multiple deployments.
- It can generate tests for edge cases, increasing the coverage and reliability of software testing processes.

Used appropriately, tools that incorporate generative AI can help developers significantly accelerate the development of experimental defense capability by enabling rapid prototyping and simulation. An ongoing challenge, however, involves defining success criteria for the many emerging uses of generative AI in software engineering [4]. Our experience at the Carnegie Mellon University (CMU) Software Engineering Institute (SEI) —a DoD federally funded research and development center (FFRDC)— indicates that integrating generative AI into the software development lifecycle (SDLC) requires a measured approach, balancing concerns like disclosure, accuracy, and ethical use [5][6][7]. Success hinges on developing organizational policies for such concerns and adapting to evolving governance and regulations.

An empirical understanding of workflow alterations and data collection helps inform decisions about the success of new approaches. For instance, by tracking the time required for automated code generation versus manual coding practices, organizations can assess productivity gains and determine the optimal integration of AI-augmented methods within their development lifecycles. Moreover, traditional practices, such as code reviews with customized checklists, may even regain prominence, providing humans in the loop with the tools and methods to ensure the reliability and testability of code and systems developed with the assistance of generative AI.



**Figure 1. Taxonomy of AI Augmentation for System Operations and the Software Development Lifecycle.**

**Figure 1** expands upon the vision presented in our 2021 book, *Architecting the Future of Software Engineering: A National Agenda for Software Engineering Research & Development*, to codify opportunities for applying AI augmentation in both system operations and the SDLC, ranging from conventional methods to fully AI-augmented methods [8][9]. Use of generative AI is a driver of the degree of AI-augmentation in the SDLC axis in the scope of our discussion in this paper, but use of AI technologies in operations or the SDLC is not limited to generative AI.

Each quadrant in **Figure 1** is summarized below:

- **Conventional systems built using conventional SDLC techniques**—This quadrant represents a low degree of AI augmentation for both system operations and the SDLC, which is the baseline of most software-reliant projects today. An example is an avionics mission computing system that uses distributed object computing middleware and rate monotonic scheduling and is developed using conventional SDLC processes without any AI-augmented tools or methods.
- **Conventional systems built using AI-augmented techniques**—This quadrant represents an emerging area of research, development, and practice in the software engineering community, where system operations have a low degree of AI augmentation, but AI-augmented tools and methods are used in the SDLC. An example is a website hosting service where the content is not AI augmented, but the development process employs AI-based code generators (such as GitHub Copilot), AI-based code review tools (such as Codiga), and/or AI-based testing tools (such as DiffBlue Cover).
- **AI-augmented systems built using conventional SDLC techniques**—This quadrant represents a high degree of AI augmentation in systems, especially in their runtime operations, but uses conventional methods in the SDLC. An example is a recommendation engine in an e-commerce platform that employs machine learning algorithms to personalize recommendations, but the software itself is developed, tested, and deployed using conventional Agile methods and the React.js and Node.js frameworks.
- **AI-augmented systems built using AI-augmented techniques**—This quadrant represents the pinnacle of AI augmentation, with a high degree of AI-augmentation for both systems operations and the SDLC. An example is a self-driving car system that uses machine learning algorithms for navigation and decision making while also using AI-driven code generators, code review and repair tools, unit test generation, and DevOps tools for software development, testing, and deployment.

Applying generative AI for AI-augmented methods in software engineering is likely to change many processes across the SDLC. Further work is needed to address potential errors unique to generative AI (e.g., new tools for detecting and addressing these errors) and methods for measuring the impact of generative AI use (e.g., on feature delivery rates and data protection). Software engineering research to date has largely focused on demonstrating application of LLMs to improve routine software engineering tasks, demonstrating improvements along building conventional systems using AI-augmented SDLC techniques. Examples include:

- Using LLMs in test automation [10]
- Converting requirements to machine readable formats [11]
- Auto code completion [12]
- Code comment generation [13]
- Program repair [14]
- Code review [15]

Works such as these, focusing on improving tasks using LLM approaches, need to be complemented by approaches which look at end to end workflows and how to complement LLMs with other automation.

Moreover, research is needed to develop specialized AI models for domains and technologies that are uncommon outside of the DoD. For example, commercial generative AI will likely favor current technologies (such as service-oriented architectures and mobile cloud computing) and popular programming languages (such as Python and Rust). Consequently, it may be hard for DoD programs to leverage generative AI capabilities in less common settings, such as maintenance and modernization of systems that use older programming languages, such as Fortran, Jovial, or even Cobol.

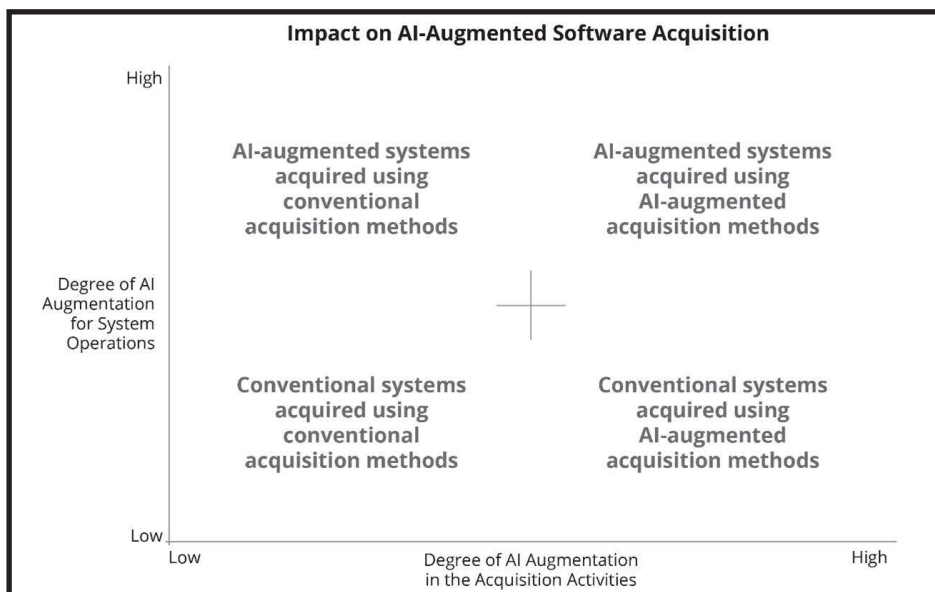
## Generative AI in Acquisition

The application of generative AI to DoD acquisition is a potentially transformative shift, offering opportunities to streamline processes, enhance strategic decision making, and optimize use of limited expertise and resources in DoD acquisition [16]. In the highly complex, heavily regulated, and security-sensitive domain of DoD acquisition, generative AI can perform several pivotal tasks, including the following:

- It can summarize voluminous policy documents (e.g., DoD directives, instructions, memoranda, and guidance) and assist in updating the documentation to increase consistency.
- It can sift through extensive regulatory policies and standards to identify the most relevant areas for a specific acquisition or system context and assist in monitoring regulatory compliance throughout the system lifecycle.
- It can assist in identifying potential risks or threats within the acquisition process, for example from cyber threats or supply chain compromises. This proactive and ongoing identification allows for the implementation of robust security measures and risk mitigation strategies.

Success in applying generative AI within defense acquisition can be evaluated via several criteria, including the enhancement of national security, the efficiency of procurement processes, the regulatory compliance of acquired defense systems, and the effectiveness of risk management. Measurable outcomes include the reduction of development and procurement timelines, improvements in the quality and performance of defense capabilities, and higher mission resilience through regulatory compliance and risk mitigation.

The application of generative AI in defense acquisition workflows similarly includes multiple risks and considerations. The reliance on generative AI to inform decision-making processes necessitates clear process and scrutiny to reduce biases and ensure data integrity. Moreover, there's the challenge of ensuring that AI-generated solutions comply with international laws and ethical standards related



**Figure 2.** Taxonomy of AI Augmentation for System Operations and Acquisition Activities.

to defense acquisition. To mitigate these risks, a well-balanced approach that combines generative AI with human expertise and oversight is crucial to ensuring defense acquisition processes remain secure, efficient, and aligned with strategic objectives.

**Figure 2** depicts opportunities for applying AI augmentation in both system operations and acquisition activities, ranging from conventional to fully AI-augmented methods. Use of generative AI is a driver of the degree of AI-augmentation in the acquisition activities axis in the scope of our discussion in this paper, but use of AI technologies in operations and the SDLC is not limited to generative AI.

Each quadrant in **Figure 2** is summarized below.

- **Conventional systems acquired using conventional acquisition methods—**  
This quadrant represents a low degree of AI augmentation (if used at all) for both system operations and acquisition, which is the baseline of the vast majority of software-reliant acquisition programs today. An example is a military-grade GPS satellite system that uses traditional data transmission and encryption for operations and is developed using conventional acquisition processes without any AI-augmented tools or methods.
- **Conventional systems acquired using AI-augmented acquisition methods—**  
This quadrant represents an emerging area of research in the acquisition community in which system operations have a low degree of AI augmentation, but AI-augmented tools and methods are used in the acquisition activities. An example is a GPS-guided munition where the content is not AI-augmented, but the acquisition activities employ AI-assistance in identifying and analyzing relevant regulations, standards, and potential security risks.
- **AI-augmented systems acquired using conventional acquisition methods—**  
This quadrant represents a high degree of AI augmentation in systems, especially in operations, but uses conventional methods in the acquisition. An example is a radar system that employs machine learning to identify and prioritize possible targets, but the system is acquired using conventional methods.
- **AI-augmented systems acquired using AI-augmented acquisition methods—**  
This quadrant represents the pinnacle of AI augmentation, with a high degree of AI-augmentation for both systems operations and the acquisition. An example is an autonomous vehicle or platform that employs AI to navigate while also using AI-augmented acquisition processes, methods, and tools, such as text summarization and semi-automated regulatory compliance.

It is important to recognize that applying advanced tool support to acquisition tasks—especially generative AI-based techniques—is still in its infancy. Further work is needed, therefore, on developing more sophisticated generative AI models that can:

1. Understand and interpret large and complex acquisition documents
2. Enhance the data analytics capabilities to forecast project outcomes and risks more accurately
3. Create more intuitive interfaces for human-AI interaction to facilitate decision making
4. Conduct comprehensive studies on the long-term impacts and ethics of AI integration into the acquisition process

Moreover, acquisition professionals must be trained to manage and collaborate with AI-augmented processes and systems effectively to enable the seamless integration of AI tools within existing acquisition workflows, as discussed in the “Essential Generative AI Skills for the Workforce of Tomorrow” section on page 36.

# Example Software Engineering and Acquisition Use Cases

The ability of LLMs to generate plausible content for text and code applications in software engineering has motivated steadily increasing experimentation by researchers and practitioners. For example, a literature review of 229 research papers written between 2017-2023 on the application of LLMs to software engineering problems finds applications spanning requirements, design, development, testing, maintenance, and management activities, with development and testing being the most common [2].

Based on our work with many government organizations, the SEI has adopted a broader perspective and formulated several dozen ideas for using LLMs in common software engineering and acquisition activities (see Table 1 for examples) [5]. Two important observations emerged from this activity. First, most use cases represent human-AI partnerships in which an LLM or generative AI service could be used to help humans complete tasks more quickly (as opposed to replacing humans). Second, deciding which use cases would be most feasible, beneficial, or affordable is a non-trivial decision for those organizations just getting started with LLMs. A discussion on developing use cases and assessing the suitability of generative AI is available in the SEI report on Assessing Opportunities for LLMs in Software Engineering and Acquisition [5].

Software Engineering Use Cases	Acquisition Use Cases
<p><b>SE1.</b> A developer uses an LLM to find vulnerabilities in existing code, hoping that the exercise will catch additional issues not already found by static analysis tools.</p>	<p><b>A1.</b> A new acquisition specialist uses an LLM to generate an overview of relevant federal regulations for an upcoming RFP review, expecting the summary to save time in background reading.</p>
<p><b>SE2.</b> A developer uses an LLM to generate code that parses structured input files and performs specified numerical analysis of its inputs, expecting it to generate code with the desired capabilities.</p>	<p><b>A2.</b> A chief engineer uses an LLM to generate a comparison of alternatives from multiple proposals, expecting it to use the budget and schedule formulas from previous similar proposal reviews and generate accurate itemized comparisons.</p>
<p><b>SE3.</b> A tester uses an LLM to create functional test cases, expecting it to produce a set of text test cases from a provided requirements document.</p>	<p><b>A3.</b> A contract specialist uses an LLM to generate ideas for an RFI solicitation given a set of concerns and a vague problem description, expecting it to generate a draft RFI that is at least 75% aligned with their needs.</p>
<p><b>SE4.</b> A developer uses an LLM to generate software documentation from code to be maintained, expecting it to summarize its functionality and interface.</p>	<p><b>A4.</b> A CTO uses an LLM to create a report summarizing all uses of digital engineering technologies in the organization based on internal documents, expecting it can quickly produce a clear summary that is at least 90% correct.</p>

**Table 1.** Sample Software Engineering and Acquisition Use Cases.



<p><b>SE5.</b> A software engineer who is unfamiliar with SQL uses an LLM to generate a SQL query from a natural language description, expecting it to generate a correct query that can be tested immediately.</p>	<p><b>A5.</b> A program office lead uses an LLM to evaluate a contractor’s code delivery for compliance with required design patterns, expecting that it will identify any instances in which the code fails to use required patterns.</p>
<p><b>SE6.</b> A software architect uses an LLM to validate whether code that is ready for deployment is consistent with the system’s architecture, expecting that it will reliably catch deviations from the intended architecture.</p>	<p><b>A6.</b> A program manager uses an LLM to summarize a set of historical artifacts from the past six months in preparation for a high visibility program review and provides specific retrieval criteria (e.g., delivery tempo, status of open defects, and schedule), expecting it to generate an accurate summary of program status that complies with the retrieval criteria.</p>
<p><b>SE7.</b> A developer uses an LLM to translate several classes from C++ to Rust, expecting that the translated code will pass the same tests and be more secure and memory safe.</p>	<p><b>A7.</b> A program manager uses an LLM to generate a revised draft of a statement of work given a short starting description and a list of concerns (e.g., cybersecurity, software delivery tempo, and interoperability goals). The program manager expects it to generate a structure that can be quickly refined and that includes topics drawn from best practices that they may not think to request explicitly.</p>
<p><b>SE8.</b> A developer uses an LLM to generate synthetic test data for a new feature being developed, expecting that it will quickly generate syntactically correct and representative data.</p>	<p><b>A8.</b> A requirements engineer uses an LLM to generate draft requirements statements for a program upgrade based on past similar capabilities, expecting them to be a good starting point.</p>
<p><b>SE9.</b> A developer provides an LLM with code that is failing in production and a description of the failures, expecting it to help the developer diagnose the root cause and propose a fix.</p>	<p><b>A9.</b> A contract officer is seeking funding to conduct research on a high priority topic they are not familiar with. The contract officer uses an LLM to create example project descriptions for their context, expecting it to produce reasonable descriptions.</p>

**Table 1.** *Sample Software Engineering and Acquisition Use Cases.*

## Deciding When (and When Not) to use Generative AI

As generative AI continues to reshape day-to-day tasks in the software engineering and acquisition ecosystems, a key question to consider is when generative AI should and should not be used. Some transformative opportunities exist that boost productivity, such as coding, testing, simulation, document analysis, and data synthesis. However, challenges like “hallucinations” (which are incorrect information generated by an LLM) and data disclosure necessitate a measured approach.

Determining the suitability of generative AI for any given task depends on assessing the nature and complexity of the task against concerns like data disclosure, accuracy, and ethical use, especially in sensitive contexts like DoD acquisition programs. Recognizing such concerns and deciding how to address each helps decision makers make more informed choices. Multiple perspectives should

therefore be considered before adopting generative AI since it sometimes produces incorrect results.

**Figure 3** depicts this perspective based on the following two questions:

- How much time and effort are needed for users to recognize that results from generative AI are incorrect?
- What are the consequences of users acting on mistaken results?



**Figure 3.** *Two Ways of Evaluating Concerns with the Generation of Incorrect Results.*

**Figure 3** shows a notional placement of the use cases from **Table 1**. The actual placement would require refinement of these use cases for specific application contexts, but the notional placement on these two questions provides insights into the opportunities for applying LLMs to a range of use cases. The upper-right (green) quadrant is ideal since mistakes have small consequences and

users can detect them with minimal effort. Use cases in this quadrant are thus a good place for organizations to begin experimenting with generative AI adoption. In contrast, the lower-left quadrant represents the least favorable use cases for applying generative AI since mistakes have large consequences and require extensive time and effort for users to detect.

Software development organizations and acquisition programs can employ several strategies to manage concerns about the use of generative AI. The following sections describe local deployment, use of domain-specific models, and the establishment of ethical use guidelines as three candidate strategies relevant to government use cases, as well as use cases for other high-stakes domains, such as healthcare, finance, and law.

## Local Deployment to Mitigate Data Disclosure Challenges

Use of commercial generative AI services often requires users to share data they operate on (e.g., prompts, code being generated, and documents being summarized) with the service provider because the models are hosted remotely. While not all use cases require sharing sensitive data, many do, which is unacceptable for defense systems, defense software engineering, and defense acquisition. For example, uploading proprietary or controlled unclassified information (CUI) documents to a generative AI service violates data disclosure rules since those documents would be ingested into the generative AI service and therefore accessible to unauthorized individuals.

Strategies for handling sensitive data disclosure in the DoD and other high-stakes domains may involve the use of synthetic data to address disclosure concerns, although this approach has limitations [17]. New approaches for security classification adherence are also needed to ensure appropriate handling of classified and unclassified data. Human oversight remains vital for task-specific

applications, with continuous human involvement ensuring that data is only disclosed as permitted by policy and regulations.

One way to significantly mitigate this risk is to rely on LLMs that are hosted on trusted networks and share no information with the model's owner. Local LLMs can include models that are trained locally, open-source models that are deployed locally, or commercial offerings that are deployed locally (e.g., complying with FedRAMP guidance). Although local LLMs may not be as powerful or up-to-date as their remote counterparts, they may be viable choices for many applications based on the success criteria, evaluation criteria, and risk concerns of each use case.

## **Use of Domain-Specific Models to Enhance Generative AI Accuracy**

Exploring the role of domain-specific models may aid the use of generative AI in specialized environments. Domain-specific LLMs are trained on data from a specific geographical or organizational context, which can capture nuances and patterns relevant to that particular environment [18]. These models contribute to improved accuracy and relevance in generating outputs tailored to local requirements, ensuring that the generated content aligns closely with the specific needs of the intended users or stakeholders. In the context of software engineering and acquisition, domain-specific models can be trained to understand and generate content that is deeply intertwined with the unique practices, terminology, and challenges of these fields.

For instance, within software engineering, domain-specific models could predict how changes to one part of a system might affect the rest or suggest software patterns that are most appropriate for a given requirement. In software acquisition, these models could simulate the project management lifecycle to forecast potential risks and outcomes, generate documentation that aligns with legal and industry standards, or optimize the allocation of resources. This tailored application of generative AI to the intricacies of software development and procurement processes can lead to more precise requirement analyses, better cost estimations, and improved strategic decision making, thereby enhancing the overall quality and reliability of software products and the efficiency of acquisition processes.

Domain-specific models for software engineering can also be trained on a vast repository of code unique to a specific programming language or framework. This specialization allows generative AI to offer more accurate and contextually relevant code suggestions, aiding developers in their coding tasks. Investing in domain-specific models for defense applications helps align their capabilities with the unique needs of hyperspectral imaging, radio frequency sensing, and other modalities.

Domain-specific models also come with challenges, however, including assembling enough quality training data and verifying output behaviors. These challenges should be evaluated carefully, so that the net benefit to the system is an improvement in productivity, capability, or some other relevant metric. Nevertheless, the incorporation of domain-specific models in generative AI has the potential to ensure a more tailored and context-aware application of AI technologies.

## **Establishing Responsible and Ethical Use Guidelines**

The responsible use of generative AI in tomorrow's workforce is critical to mitigate the potential risks and ethical concerns associated with this technology. Responsible and ethical development and use of generative AI is an area of concern that spans multiple activities, including (but not limited to) data collection and preparation, model development, and use of generative AI tools and services. Challenges include the perpetuation of biases inherent in training data, the necessity for consistent monitoring and updates to prevent misuse, and the complexities surrounding the explainability of

sophisticated models. Addressing these challenges requires an ongoing commitment to research, collaboration, and transparency to foster an equilibrium between innovation, development, and responsible use. It is critical to establish the scope of activities for intended use and clarify the guidelines that are most applicable to stakeholders.

The capabilities of generative AI models are evolving rapidly, so it is essential to educate users on their responsible use. There is an emerging consensus among developers and users that the most effective generative AI tools are those that empower users with control over data privacy, model training parameters, and content generation constraints [19]. Usage patterns across software engineering and acquisition reveal a consistent interactive cycle that includes prompting the AI, executing an action based on its response, and then proceeding with further prompts.

AI-augmented methods should keep humans in the loop for multiple reasons, one of which is to be a safeguard and take responsibility for the outcome. Generative AI does make mistakes, so humans should operate with that assumption and compensate. This iterative, human-in-the-loop approach underscores the critical need for guidelines on the appropriate use of generative AI, accentuating the pivotal role of users. Some guidelines will be straightforward (such as simply reminding users of the organization's data privacy and information disclosure policies) whereas others may require strategies that include limiting use of generative AI services for particular use cases.

The landscape of responsible and ethical AI development and application is expanding, with numerous frameworks emerging to mitigate AI's unintended impacts, including those from generative AI. For instance, the Defense Innovation Unit (DIU) has formulated Responsible AI (RAI) guidelines to streamline the evaluation process for those involved in AI project development, such as program managers, commercial vendors, or government collaborators [20]. These guidelines cover a broad spectrum of considerations, including legal, procurement, technical, and operational aspects. They offer directives for both AI tool developers and users; for instance, outlining the extent of technical transparency required while safeguarding proprietary data. Organizations are encouraged to augment these guidelines with their specific procedures and data-sharing policies, ensuring alignment with their domain-specific requirements.

The broad adoption of responsible AI in software engineering and acquisition is also contingent on legal maturity. As service providers begin to indemnify outputs from generative AI tools against intellectual property infringements, we anticipate the formation of a trusted ecosystem. This ecosystem will be critical in fostering responsible use and ensuring that generative AI is leveraged to enhance, rather than compromise, the quality and reliability of software products and services.

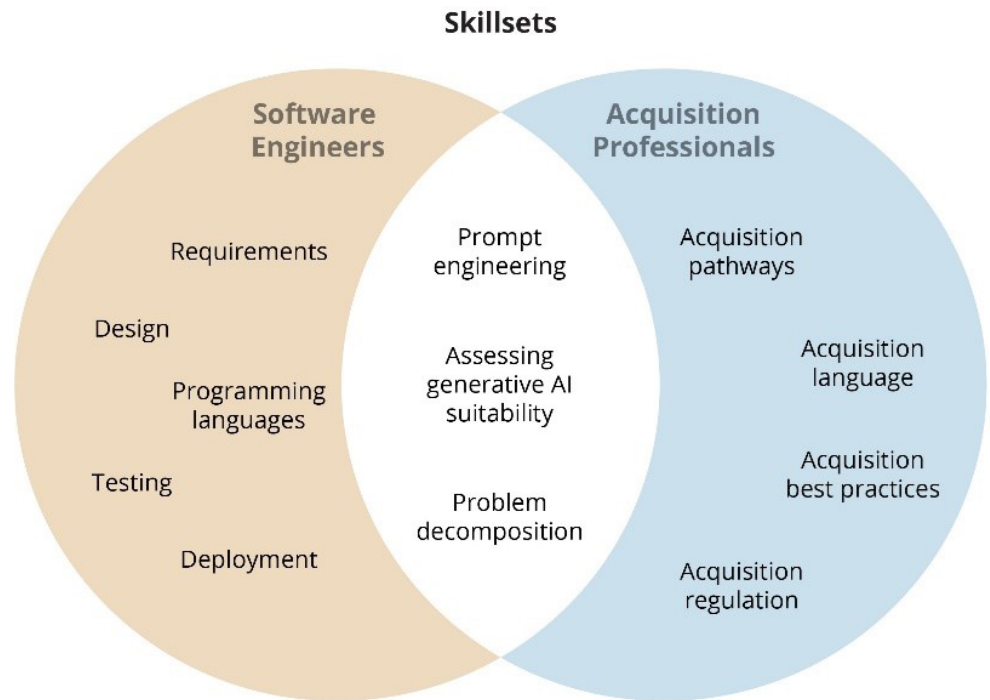
## Essential Generative AI Skills for the Workforce of Tomorrow

Generative AI will augment, rather than replace, the capabilities of software engineers and acquisition professionals for the foreseeable future. Consequently, workers in both professions must maintain expertise in their respective domains. Software engineers will need proficiency with requirement analysis, software design, programming languages, testing practices, and deployment. Likewise, acquisition professionals will need proficiency with acquisition regulations, acquisition pathways, and their application to different system contexts [21].

To unlock the potential of generative AI, however, software engineers and acquisition professionals should cultivate new skills, such as those visualized in **Figure 4**. For example, both should learn

prompt engineering and how to apply prompt patterns to elicit effective results from generative AI [22]. Likewise, both should master decomposing complex issues into manageable components and assessing which of these issues generative AI can help solve. Above all, users who responsibly exercise curiosity, experimentation, and a willingness to learn new skills will guide the DoD in successfully adapting to the dynamic landscape of generative AI.

When skillsets are visually summarized (as seen in **Figure 4**), it becomes clear that generative AI does not replace software and acquisition professionals, but rather augments their effectiveness through new skills, such as prompt engineering and problem decomposition. Individual effectiveness in utilizing generative AI may vary based on skills, experience, and adaptability. Some individuals may naturally excel in leveraging these tools for code generation, problem-solving, and documentation, whereas others may require more extensive training. Either way, continuous learning and adaptability are key.



**Figure 4.** *Representative Skillsets for Software Engineers and Acquisition Professionals Using Generative AI.*

Given generative AI's tendency to make mistakes, validating the outputs of generative AI is an essential activity for both the software engineering and acquisition communities. The specific skills needed to validate output will vary with nature of the task and data, but two questions should always be considered: Is the information in the output correct, and is any information missing from the output? Here are some examples:

- **Generating source code for specific requirement**—Users can write unit tests to confirm that computations performed by generated code are correct. Likewise, users can inspect the generated code to confirm that it does not performing any unnecessary work.
- **Summarizing document contents from a specific stakeholder perspective**—Users can request a summary of main points from a document relevant for a specific stakeholder (e.g., a software safety engineer, reliability engineer, cybersecurity analyst, etc.) and fact check the summary by searching the source material for relevant facts. Users can review the source material to confirm that essential and relevant points are included in the summary.

Incorporating generative AI tools into the software engineering educational curricula will help the emerging workforce [23]. Of course, existing software engineers must stay abreast of advancements throughout their career because changes happen quickly. Generative AI tools can also assist the acquisition workforce in keeping current with updates to acquisition regulations. However, acquisition professionals are responsible for codifying significant shifts in best practices and new system types, pending the availability of adequate new data to train generative AI tools.

Organizations offering training for acquisition professionals, such as Defense Acquisition University, will also need to incorporate generative AI into their curricula. While generative AI itself may not replace jobs, those who excel in using generative AI tools might surpass others in the job market. As the landscape of system development evolves, adapting to change and acquiring skills in the use of generative AI will be crucial for staying competitive in the dynamic and exciting future envisioned by the software engineering community.

## **Conclusion: We Must Learn to Navigate an AI-augmented Future for Software Engineering and Acquisition**

The initial adoption phase of generative AI in software engineering and acquisition will likely be tumultuous as users navigate the applicability and utility of these tools to different tasks. Some ideas will be highly successful, whereas others will prove disappointing. This exploratory phase is crucial as we collectively learn the potential of generative AI to shape future research and application throughout the DoD.

Software engineers across the globe are already applying generative AI in software engineering today, demonstrating practical applications in code generation for routine tasks. This early adoption has the potential to grow into more impactful applications, including accelerating software modernization (e.g., by resolving technical debt and repairing critical errors) and quickly assembling prototypes (e.g., by crawling software repositories to identify compatible candidate software elements). Similarly, application of generative AI to acquisition activities has potential to improve the efficiency of summarization and document generation of acquisition artifacts.

Generative AI significantly lowers the barrier to entry for content generation, with potentially mixed implications. These technologies are empowering users without formal software engineering backgrounds to solve complex problems using natural language interfaces, which opens access to the ideas and imagination of a much larger population. This empowerment also brings challenges, however, especially in terms of ensuring the quality of generated content when users lack the deep technical knowledge traditionally associated with software engineering roles. The impact of incorporating generated content without the benefit of conventional engineering review on system stability, security, and operational accuracy are unknown. It is important to recognize that generative AI services are tools to assist users, they do not replace expertise in software engineering and acquisition.

The educational landscape for software engineering and acquisition must evolve to integrate generative AI, preparing students and professionals alike to harness these tools effectively while also understanding their limitations and inherent biases. This curriculum development will facilitate a new breed of software engineers and acquisition professionals skilled in generative AI use and critical evaluation, ensuring their ability to innovate responsibly while maintaining ethical standards. These future-focused educational strategies are essential as generative AI increasingly becomes integral to many disciplines and domains, emphasizing a blend of technical proficiency with a thorough grasp of AI's ethical and practical implications. This balanced approach will foster professionals who can navigate the complexities of using generative AI and contribute to its ethical advancement.

In conclusion, navigating the future AI-augmented software engineering and acquisition is a tapestry of opportunity and responsibility, weaving together advancements in AI with the need for ethical stewardship and thoughtful integration into high-stakes DoD socio-technical systems.

# Acknowledgements

We would like to thank Erin Harper and Ed Desautels for editorial and design support.

## Copyright

Copyright 2024 Carnegie Mellon University.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

References herein to any specific entity, product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by Carnegie Mellon University or its Software Engineering Institute nor of Carnegie Mellon University - Software Engineering Institute by any such named or represented entity.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN “AS-IS” BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. Requests for permission for non-licensed uses should be directed to the Software Engineering Institute at [permission@sei.cmu.edu](mailto:permission@sei.cmu.edu).

Carnegie Mellon® is registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM24-0335

## References

- [1] McKinsey & Company. “What is Generative AI?” McKinsey & Company Featured Insights. 19 Jan. 2023. <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai/#/>. Accessed 19 Mar. 2024.
- [2] Hou, X. et al. “Large Language Models for Software Engineering: A Systematic Literature Review.” arXiv, 12 Sept. 2023. <https://arxiv.org/pdf/2308.10620.pdf>.
- [3] Jiang, Ellen, et al. “PromptMaker: Prompt-Based Prototyping with Large Language Models.” Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems - CrossTalk - May 2024

tems, Association for Computing Machinery, 2022, pp. 1–8. ACM Digital Library, <https://doi.org/10.1145/3491101.3503564>.

[4] A. Fan et al., “Large Language Models for Software Engineering: Survey and Open Problems,” 2023 IEEE/ACM International Conference on Software Engineering: Future of Software Engineering (ICSE-FoSE), Melbourne, Australia, 2023, pp. 31-53, doi: 10.1109/ICSE-FoSE59343.2023.00008.

[5] Ozkaya, Ipek. “What Is Really Different in Engineering AI-Enabled Systems?” IEEE Software, vol. 37, no. 4, pp. 3–6, July-Aug. 2020, <https://doi.org/10.1109/MS.2020.2993662>.

[6] Bellomo, Stephany et al. “Assessing Opportunities for LLMs in Software Engineering and Acquisition.” Software Engineering Institute, 1 Nov. 2023, <https://insights.sei.cmu.edu/library/assessing-opportunities-for-llms-in-software-engineering-and-acquisition/>.

[7] Schmidt, Douglas et al. The Future of Software Engineering and Acquisition with Generative AI. Software Engineering Institute, 23 Jan. 2024, <https://insights.sei.cmu.edu/library/the-future-of-software-engineering-and-acquisition-with-generative-ai/>.

[8] Carleton, Anita et al. Architecting the Future of Software Engineering: A National Agenda for Software Engineering Research & Development. Software Engineering Institute, 2021, <https://insights.sei.cmu.edu/library/architecting-the-future-of-software-engineering-a-national-agenda-for-software-engineering-research-development/>.

[9] Ozkaya, Ipek et al. “Application of Large Language Models (LLMs) in Software Engineering: Overblown Hype or Disruptive Change?” SEI Blog, 2 Oct. 2023, <https://insights.sei.cmu.edu/blog/application-of-large-language-models-llms-in-software-engineering-overblown-hype-or-disruptive-change/>.

[10] Z. Liu, C. Chen, J. Wang, M. Chen, B. Wu, X. Che, D. Wang, Q. Wang. “Make LLM a Testing Expert: Bringing Human-like Interaction to Mobile GUI Testing via Functionality-aware Decisions.” 46th IEEE/ACM ICSE 2024.

[11] Tikayat Ray, Archana, et al. “Agile Methodology for the Standardization of Engineering Requirements Using Large Language Models.” Systems, vol. 11, July 2023, p. 352. <https://doi.org/10.3390/systems11070352>.

[12] Bird, Christian, et al. “Taking Flight with Copilot: Early Insights and Opportunities of AI-Powered Pair-Programming Tools.” Queue, vol. 20, no. 6, Jan. 2023, p. Pages 10:35-Pages 10:57. November/December, <https://doi.org/10.1145/3582083>.

[13] Geng, Mingyang, et al. “Large Language Models Are Few-Shot Summarizers: Multi-Intent Comment Generation via In-Context Learning.” IEEE Computer Society, 2024, pp. 453–65. [www.computer.org](http://www.computer.org), <https://www.computer.org/csdl/proceedings-article/icse/2024/021700a453/1WDJaRUZRKg>.

[14] Jin, Matthew, et al. “InferFix: End-to-End Program Repair with LLMs.” Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Association for Computing Machinery, 2023, pp. 1646–56. ACM Digital Library, <https://doi.org/10.1145/3611643.3613892>.

[15] Li, Lingwei, et al. “AUGER: Automatically Generating Review Comments with Pre-Training Models.” Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Association for Computing Machinery, 2022, pp. 1009–21. ACM Digital Library, <https://doi.org/10.1145/3540250.3549099>.

[16] Robert, John, and Douglas Schmidt. “10 Benefits and 10 Challenges of Applying Large Language Models to DoD Software Acquisition.” SEI Blog, 22 Jan. 2024, <https://insights.sei.cmu.edu/blog/10->



benefits-and-10-challenges-of-applying-large-language-models-to-dod-software-acquisition/.

[17] Li, Zhuoyan, et al. "Synthetic Data Generation with Large Language Models for Text Classification: Potential and Limitations." arXiv:2310.07849, 12 Oct. 2023. arXiv.org, <https://doi.org/10.48550/arXiv.2310.07849>.

[18] Nucci, Antonio. "What Is a Domain-Specific LLM? Examples and Benefits." Aisera: Best Generative AI Platform For Enterprise, 11 Jan. 2024, <https://aisera.com/blog/domain-specific-llm/>.

[19] Zhang, Dawen et al. "Privacy and Copyright Protection in Generative AI: A Lifecycle Perspective." 2024 IEEE/ACM 3rd International Conference on AI Engineering - Software Engineering for AI (CAIN), 2024.

[20] Defense Innovation Unit. "Responsible Artificial Intelligence: 2022 in Review." Defense Innovation Unit Artificial Intelligence Portfolio, 1 May 2023, <https://www.diu.mil/responsible-ai-guidelines>. Accessed 19 Mar. 2024.

[21] Defense Acquisition University. "Adaptive Acquisition Framework." DAU, <https://aaf.dau.edu>. Accessed 19 Mar. 2024.

[22] White, Jules et al. "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT." Proceedings of the 30th Pattern Languages of Programming (PLoP) Conference, Allerton Park, IL, October 23-25th, 2023.

[23] Kirova, Vassilka D. et al. "Software Engineering Education Must Adapt and Evolve for an LLM (Large Language Model) Environment." Proceedings of the 55th ACM Technical Symposium on Computer Science Education, March 2024.

## About the Author



John E. Robert is a Principal Engineer at the SEI and the Deputy Director for the SEI's Software Solutions Division. Robert provides leadership for software engineering research and the development of technologies in partnership with DoD programs and industry to enable the broad transition of new software engineering approaches. Robert has led multiple SEI technical partnerships with high-priority DoD programs, and was part of the lead author team on a national study for software engineering research and development in 2021 titled, *Architecting the Future of Software Engineering: A National Agenda for Software Engineering Research & Development*.

**John E. Robert**

**Deputy Director**

**Software Engineering Institute**

**[jer@sei.cmu.edu](mailto:jer@sei.cmu.edu)**



James Ivers is a Principal Engineer and lead of the Architecture Design, Analysis, and Automation group at the SEI. His experience spans research and application of work in software architecture, code analysis, formal methods, and scaling our ability to evolve software. His most recent work focuses on using artificial intelligence (AI) for software engineering to support large-scale refactoring.

**James Ivers**

**Principal Engineer**

**Software Engineering Institute**

**[jivers@sei.cmu.edu](mailto:jivers@sei.cmu.edu)**



Dr. Doug Schmidt is the Cornelius Vanderbilt Professor of Engineering, Associate Chair of Computer Science, and a Senior Researcher at the Institute for Software Integrated Systems, all at Vanderbilt University. He is also a Visiting Scientist at the Software Engineering Institute at Carnegie Mellon University. Dr. Schmidt is an internationally renowned and widely cited researcher whose work focuses on pattern-oriented middleware, Java concurrency and parallelism, and generative AI.

**Dr. Doug Schmidt**

**Professor of Engineering**

**Vanderbilt University**

**[d.schmidt@vanderbilt.edu](mailto:d.schmidt@vanderbilt.edu)**



Dr. Ipek Ozkaya is a principal researcher and the technical director of the Engineering Intelligent Software Systems group at the Software Engineering Institute. Her areas of work include software architecture, software design automation, and managing technical debt in software-reliant and AI-enabled systems. At the SEI, she has worked with several government and industry organizations in domains including avionics, power and automation, IoT, healthcare, and IT. Ozkaya is the co-author of a practitioner book titled *Managing Technical Debt: Reducing Friction in Software Development*.

**Dr. Ipek Ozkaya**

**Technical Director**

**Software Engineering Institute**

**ozkaya@sei.cmu.edu**



Shen Zhang is a Senior Software Engineer within the Enabling Mission Capability at Scale directorate at the SEI. He provides expertise in software development processes and static code analysis for real-time mission critical software-intensive systems in the Air Force and Navy. More recently, he has been involved with using large language models to accelerate software engineering use cases. Prior to working at the SEI, he developed simulation products for industrial control systems in the commercial power industry.

**Shen Zhang**

**Software Engineer**

**Software Engineering Institute**

**szhang@sei.cmu.edu**

# SAVE THE DATE!

## 2024 DoD Weapon Systems Software Summit

**December 10-12, 2024  
Salt Lake City, UT  
Salt Palace Convention Center**

**Keynotes, panel discussions and technical presentations from DoD and DIB colleagues on their solutions to common software problems.**



**Contact us:**

[AFSC.DoD.WS\\_SWSummit@us.af.mil](mailto:AFSC.DoD.WS_SWSummit@us.af.mil)

# U.S. Air Force Depot Operating Efficiency and Mission Readiness Using Artificial Intelligence/Machine Learning (AI/ML)

**Abeezar Tyebji**  
**Chief Executive Officer**  
**Shipcom**

**Vik Chauhan**  
**Federal Practice Lead**  
**Shipcom**

## Introduction

The United States Air Force (USAF) faces increasing pressure to extend the life of its aircraft beyond their intended designs, placing strains on aircraft maintenance operations. Aircraft age accelerates deterioration of individual subsystems, irrespective of flight use, introducing further workload complexity and volume to maintenance operations. Extending the operational lifetime of aircraft, with its increasing rate of subsystem deterioration, typically exceeds planned-out year resources, bringing further stressors “to do more with less.”

While the USAF has established integrity programs to manage aircraft subsystems, these programs largely operate in silos with little ability to create cross-sectional views across multiple subsystems. Further challenging the USAF, these programs rely on complex, pre-designed work patterns and purpose-built data-systems, making it nearly impossible to meet challenging risks. Lastly, USAF lacks Anomaly and Outlier Detection with the data scattered in the various silos.

In response to these challenges, this article proposes utilizing breakthrough data sci-

**“We need to rebuild [the science and engineering expertise and operational analysis capabilities that have deteriorated in the Air Force] . . . recruiting and growing people and giving them experiences that make them more technologically capable, and in a better position to make technological judgments about what technologies are ready and what can be done with them.”**

**-Air Force Secretary  
Frank Kendall [1]**

ence technologies to support the following overarching goals for the C-5, C-17, C-130, F-15, and A-10 aircraft programs:

- Institute Ongoing Practice of Anomaly and Outlier Detection
- Improve Depot Induction Decision-Making
- Provide Command View of Weapon Systems Readiness
- Establish Data-Science Culture for USAF Personnel

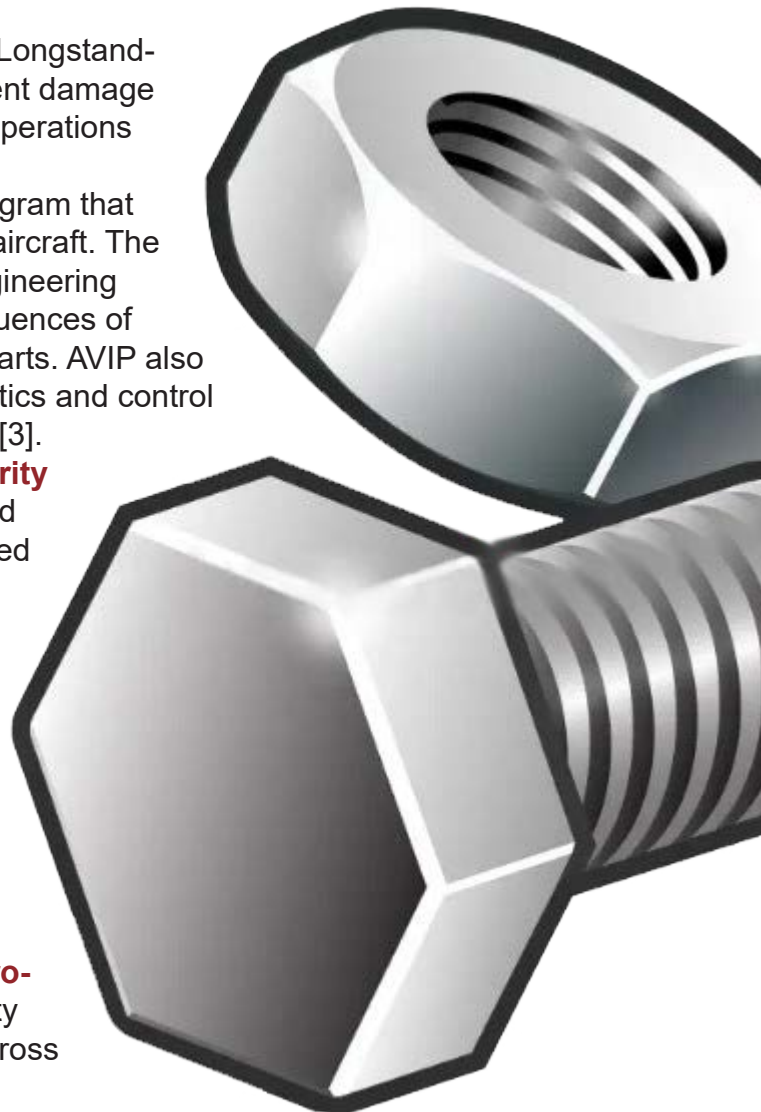
The concepts and use cases covered in the article are vetted with AFSC, NWC, AFRL and engineers including Smart Manufacturing Innovation Lead Mr. Frank Zahiri at Warner Robins Air Logistics Complex.

**“AI has been a critical tool to help increase weapon systems readiness throughout its life cycle.”**

**-Frank Zahiri, Smart Manufacturing Innovation Lead, Tech POC of Shipcom AI/ML project AFSC, Robins AFB, GA [2]**

The USAF employs numerous integrity programs to ensure the airworthiness of aircraft:

- **Aircraft Structural Integrity Program (ASIP):** Longstanding program to identify, track, resolve, and prevent damage throughout distributed and depot maintenance operations and inspections.
- **Avionics Integrity Program (AVIP):** Newer program that focuses on electronics control systems used in aircraft. The program “employs...physics, chemistry, and engineering principles to ensure an understanding of the influences of the usage and environments on materials and parts. AVIP also focuses on key product and process characteristics and control of variability of materials, parts, and processes” [3].
- **Mechanical Equipment and Subsystem Integrity Program (MECSIP):** Program using time phased actions, procedures, analyses, tests, etc. intended to ensure reliable, affordable, and supportable aircraft equipment and subsystems, thus contributing to the enhancement of total systems mission effectiveness and operational suitability.... Applies to subsystems and equipment whose operation is primarily electrical or mechanical (e.g. environmental control, fuel, flight controls, auxiliary power, electric power and wire, hydraulic systems, wheels, tires and brakes, auxiliary power, etc.).
- **Computer Systems and Software Integrity Program (CSSIP):** Newer program targeting integrity of the multiple software systems that operate across numerous aircraft programs.



Though not without its own challenges, ASIP is far and away the most established of these programs, leveraging principles from commercial aircraft maintenance and integrity monitoring. Each of these monitoring efforts is focused on detecting problems within their prescribed boundaries as defined by the data used to inform a fixed set of algorithms.

## Current State of Maintenance

Two predominant approaches are used to govern USAF maintenance:

- **Programmed Depot Maintenance (PDM):** Fixed interval, periodic inspection and correction of defects that require scope of tools, infrastructure, and skillsets not available at operating locations. High cost and time-consuming.
- **Condition Based Maintenance (CBM):** Derived from real-time, embedded sensor-based, or external sensor technologies. Maintenance is predicted when one or more indicators are at identified threshold values.

Due to the lack of accurate, multi-dimensional, predictive tools and the difficulty assembling and adjusting data feeds to support different analyses, CBM has had difficulty replacing the much more costly and much more inefficient PDM approach. What is required to support readiness is the ability to deploy accurate, sensitive, and evolving sensing systems to detect patterns of indicators that are predictive of individual and multiple component failures.

## Moving Beyond Adequacy

The current integrity programs are adequate, but they are not sufficient for the stressors of aircraft fleets aging beyond their design lives. Further, the current approaches to maintenance, favoring PDM, will place greater shares of aging fleets out of effective mission availability as the frequencies of PDM cycles increase.

The current approaches, with their intrinsic rigidity (e.g. integrity programs focused on one specific type of system) will fail to identify interrelationships between aircraft systems. It is in these synapses where none of the current integrity programs are looking that “the red needle” in the haystack awaits – the unpredicted set of variables that eluded a fixed set of detection methods.

## A New Perspective is Required

The current integrity programs and current PDM frameworks should continue. They serve their purposes well. However, leveraging advances in data science, data fabric, and Artificial Intelligence and Machine Learning, Intelligent Integrity Layers and Intelligent Lifecycle Management will both bolster current programs and mitigate gaps in detection and prediction.

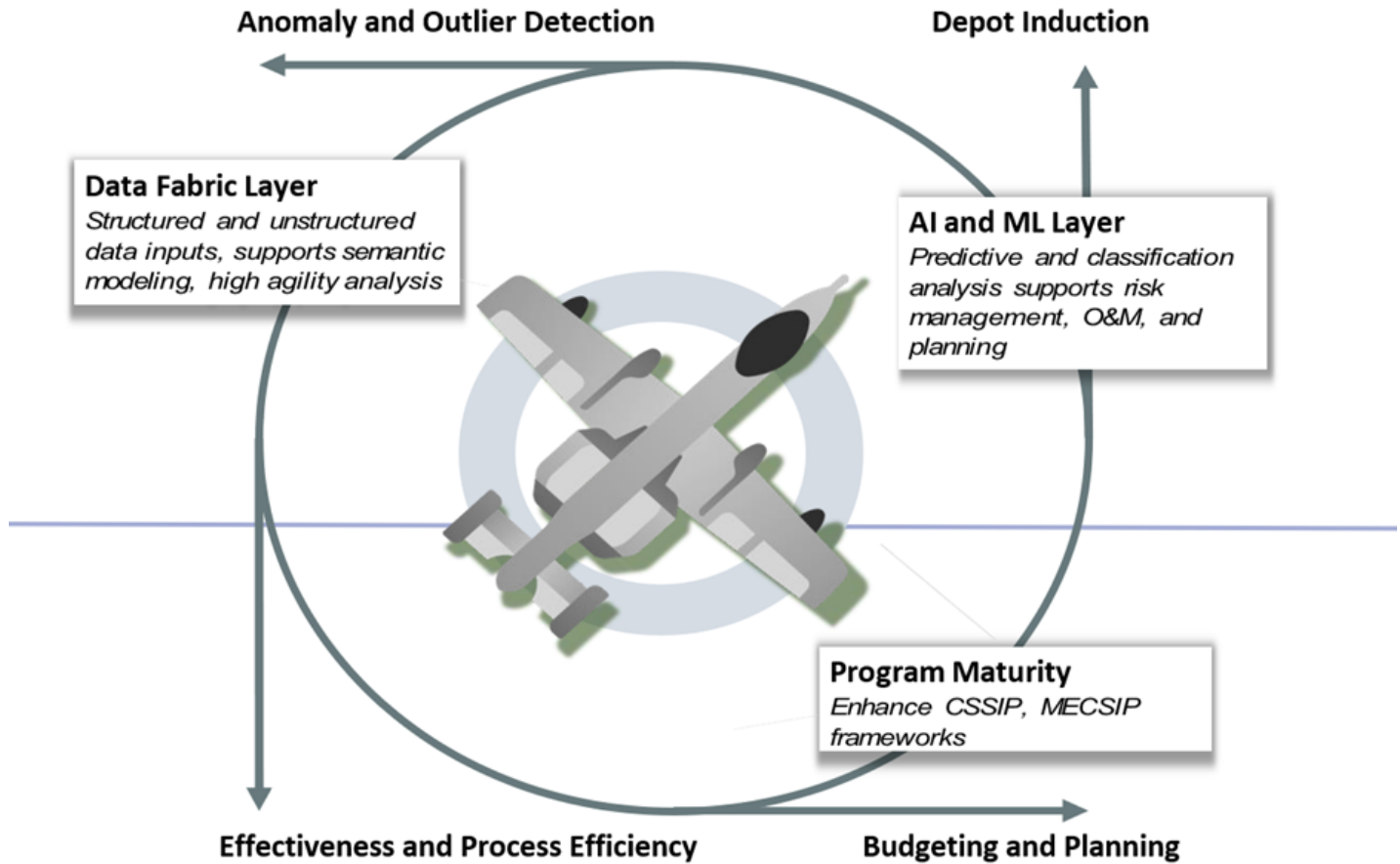
**Figure 1** provides a conceptual view of Intelligent Layers. In the figure, current Integrity Programs and certain depot processes are also depicted.

Additional intelligent layers are provided to enhance current data-driven programs. ASIP, AVIP, MECSIP, and CSSIP apply their current approaches to monitoring aircraft and their subsystems.

**Data Fabric:** The key to agility in pattern detection is unfettered and rapid access to data. Without this underlying flow of information, it will not be possible to introduce higher orders of analysis, classification, and prediction. The proposed data fabric layer is introduced to more easily facilitates adding

new types of information to current integrity programs and enables synthesis of data across multiple integrity programs. Semantic models enable the versatility of data ingestion.

**Anomaly and Outlier Detection:** The AI/ML layer provides multiple forms of time series, unsupervised classification, prediction, and optimization modeling. This versatile analytical layer supports identification of anomalies and unclassified risks across integrity programs and supports optimization to optimize operations.



**Figure 1.** AI/ML Layers to Enhance Integrity and Lifecycle Operations.

**Anomaly Detection and Composite Monitoring:** AI/ML techniques additionally enable introduction of powerful analytical methods such as “weak signals” analysis and visualizations leveraging quantum graphs and semantic radar. These approaches will enable USAF analysts to develop the ability to identify anomalous patterns that are not captured by current integrity programs. Through the use of AI/ML techniques, the USAF will be able to establish composite monitoring programs that supplement the current cadre of integrity monitoring programs.

**Program Maturity:** Though not the driving force, the adoption of Intelligent Layers powered by Data Fabric, AI/ML methods would enhance program maturity. The proposed approach provides the opportunity to enable maturation of current integrity programs. For instance, CSSIP can be enhanced to resemble the National Software Reference Library (NSRL) more closely.

**Depot Induction:** Taken together, AI/ML, and data fabric can support a better integration of CBM and PDM approaches, enabling USAF to better model and predict consumption of components consumed through maintenance work orders.

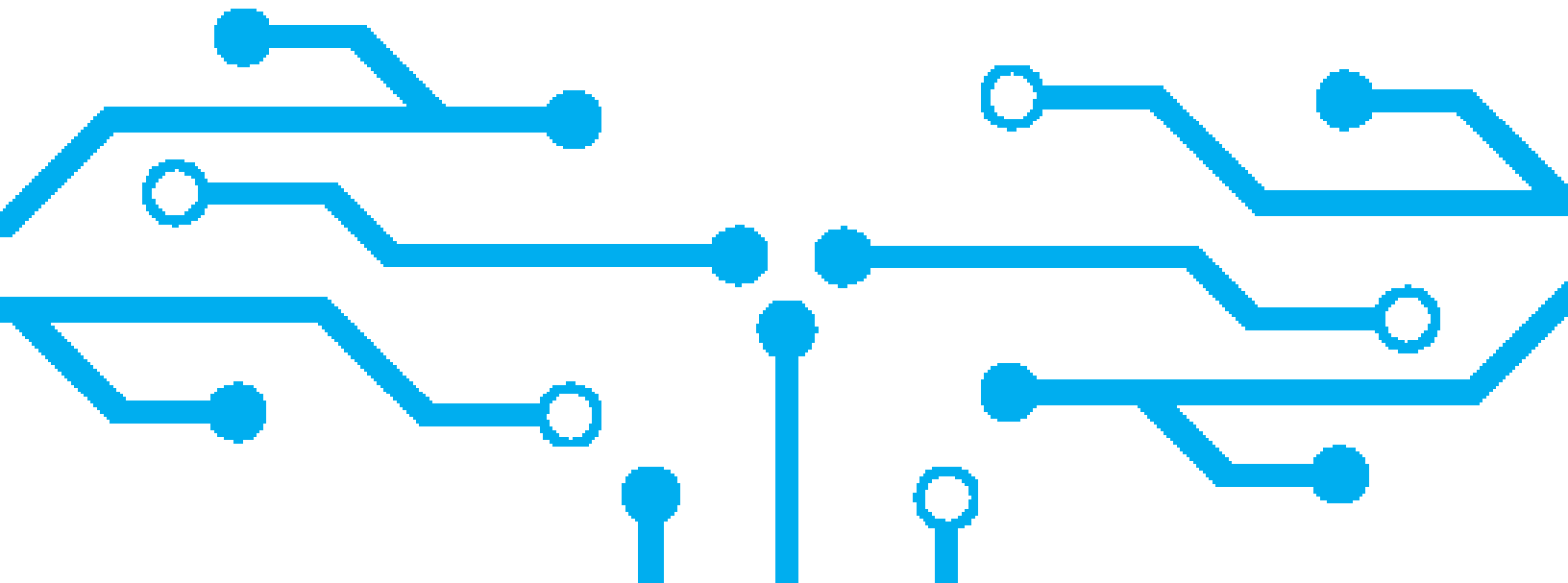


# Benefits of Intelligent Layers of Data Fabric and AI/ML

The benefits to the USAF maintenance programs of the proposed Intelligent Layers are summarized in **Table 1**.

Benefit	Nature of Benefit	Discussion
Early identification of anomalous and outlier patterns	Risk Management and Mitigation	As air fleets age beyond their design lives, it is necessary to establish the capacity to identify and manage predictors of component failures that are not captured by current analytical approaches.
Improve depot induction decision-making	Performance Improvement	AI/ML and data fabric enable complex models to better predict components to support both routine maintenance and condition-based maintenance. This approach improves the efficiency of depot programs and enables greater availability of aircraft.
Accelerate maturation of Key programs	Risk Management and Mitigation	AI/ML and data fabric can be applied to existing integrity programs to improve maturity through extensible data models and complex modeling.
Adopt data-science culture for personnel	Personnel Development	USAF military and civilian staff are more completely enculturated into using data-science and technology to better manage lifecycle operations.

**Table 1.** *Benefits of Intelligent Layers.*



# Program Scope and Approach

It is recommended that a pilot program be established whose primary mission is to accomplish the following:

- Implement Data Fabric: Establish semantic models that allow for anomaly detection between and beyond the current integrity programs.
- Anomaly and Outlier Detection: Establish analytical frameworks that support the ongoing detection of anomalous and/or outlier patterns. Establish follow-through methodologies to classify these values according to rigorous risk management framework.
- Adopt Data-Science Culture: Provide knowledge-transfer and training to USAF military and civilian personnel in the use and extension of key data-science technologies and methods.

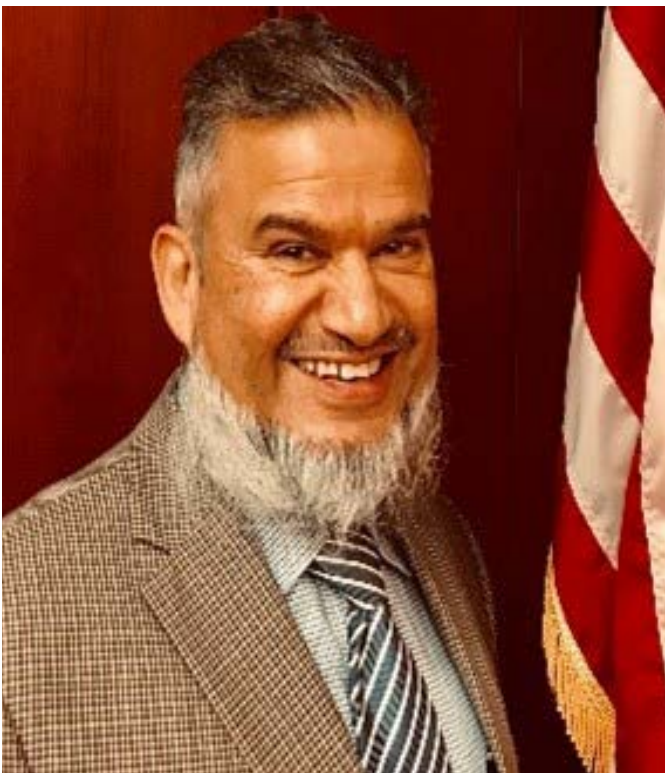
## Resources

[1] Tirpak, John A., “Mid-Tier Programs Running Out of Time; Overruns Coming, Kendall Says”, *Air and Space Forces Magazine*, *Air and Space Forces Association*, Arlington, VA, July 27, 2022.

[2] Frank Zahiri, Smart Manufacturing Innovation Lead, Tech POC of Shipcom AI/ML project AFSC, Robins AFB, GA.

[3] United States Air Force Science Advisory Board, Sustaining Air Force Aging Aircraft into the 21st Century, Internal Document, Washington, DC, August 1, 2011.

## About the Authors



Abeezar Tyebeji as the Chief Executive Officer of Shipcom, leads a team of innovators and experts in delivering AI based software solutions that leverage the power of IoT, AI, data fabric, and cyber security.

**Abeezar Tyebeji**

**Chief Executive Officer**

**Shipcom**

**[atyebji@shipcomwireless.com](mailto:atyebji@shipcomwireless.com)**



Vik Chauhan leads the Federal Practice of Shipcom AI. Shipcom AI is EASY/RESPONSIBLE AI, which is a paradigm shift from command level AI. It is an end-to-end Responsible / Ethical AI platform with data engineering and machine learning platform for engineers, scientists, analysts, and DevOps to streamline ML solution development, deployment, management, monitoring, and governance.

**Vik Chauhan**

**Vice President, Defense Programs**

**Shipcom**

**[vchauhan@shipcomwireless.com](mailto:vchauhan@shipcomwireless.com)**

# NOW HIRING

## 309TH SOFTWARE ENGINEERING GROUP, HILL AFB

- Software Engineers
- Electrical Engineers
- Aerospace Engineers
- Mechanical Engineers
- Physicists
- Systems Engineers
- IT/Cybersecurity Specialists

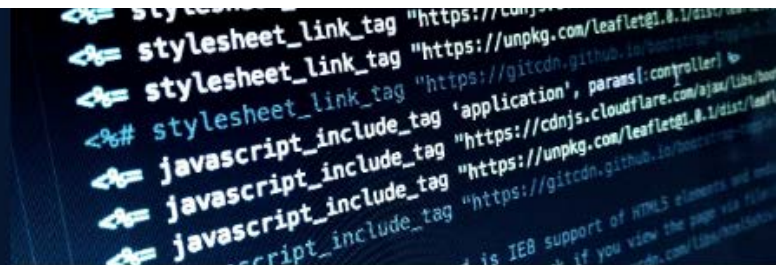
Learn what it means to **CODE WITH HONOR**

**APPLY TODAY**



**MORE INFO**

**[309smxg.recruiting@us.af.mil](mailto:309smxg.recruiting@us.af.mil)**



# The 402d Software Engineering Group is hiring!

**Location:** 402 SWEG @ Robins AFB, GA

- Electronics Engineers (0855)
- Computer Engineers (0855)
- Computer Scientists (1550)
- IT/Cybersecurity Specialists (2210)

***To apply, submit your resumes to:***  
**[AFSC.SWH.HumanCapital@us.af.mil](mailto:AFSC.SWH.HumanCapital@us.af.mil)**



**US CITIZENSHIP REQUIRED**

# IMPERFECT MATES: HUMANS AND AI IN THE COCKPIT

**Maj. Richard C.  
Agbeyibor**  
**Flight Test Engineer,  
United States Air Force**

**Dr. Karen M. Feigh**  
**Professor,  
Georgia Institute of  
Technology**

## Abstract

Rapid developments in Artificial Intelligence (AI) are bringing increasingly complex autonomy capabilities to the cockpit. Autonomous electric Vertical Take Off and Landing aircraft, swarms, Collaborative Combat Aircraft, and other new aviation mission constructs are on the horizon. In the last few decades, military and civil aviation have achieved remarkable safety and effectiveness thanks to automation and a deliberate focus on teamwork. As automation gets replaced by autonomy, the challenges of automation could be exacerbated. Effective Human-AI teaming requires both collaborative task work and teamwork which will be critical for continued safety and mission effectiveness. Despite the incredible ability of expert operators to make exceptional judgment calls in highly stressful situations, humans suffer from cognitive biases that may pose a challenge to this teaming. AI brings incredible data processing capabilities to the team but can suffer from a lack of adaptability to its environment and teammates, particularly in collaborative settings. As pilots retrain Crew Resource Management for their new AI mates, AI will also need to learn to adapt to its human mates. System developers can help achieve effective human-AI teaming by providing bidirectional transparency through interface design and system features such as status, feedback, planning mechanisms, and engagement prompts.

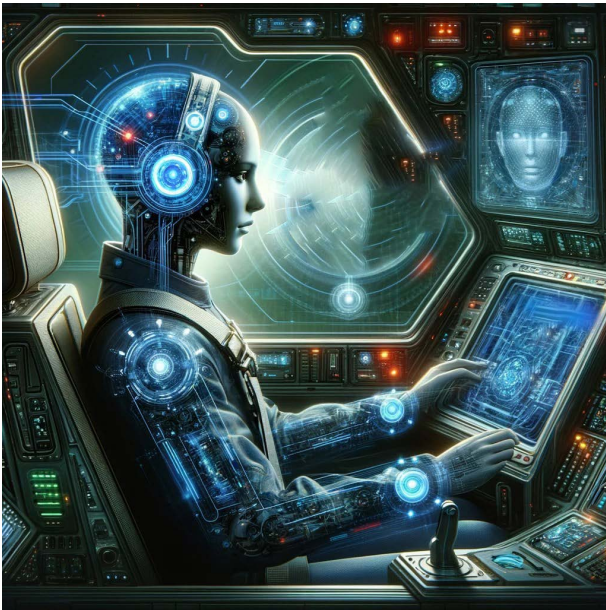
## Introduction

AI is coming and it is going to be enmeshed in every part of Air Force operations, including inside the cockpit. The field of Artificial Intelligence (AI) has gone through several booms and busts since it first surged in the 1950s and weathered the first AI Winter in the 1970s. Over the decades since, AI has benefited from exponential growth in computational power and storage over the decades, as predicted by Moore's law. Developments in algorithm design, computer engineering, networking and other related fields have led to cloud computing. This and sustained research in machine learning, reinforcement learning, natural language processing, and computer vision have enabled the latest boom that has brought Deep Learning AI to everyday consumers.

The Air Force is investing heavily in the development of AI-enabled aircraft. In his keynote speech at the 2023 Air Force Association (AFA) Symposium, the Secretary of the Air Force unveiled a plan to build and deploy a thousand Collaborative Combat Aircraft (CCA) in the near future [1]. These AI-enabled CCAs will be designed to team with fifth and sixth-generation fighter aircraft and perform a range of different missions – from carrying electronic warfare pods to forward deploying weapons and other sensors. Beyond the immediate vision of CCA, the core enabling autonomy technologies will usher in the realization of additional human-AI missions such as autonomous refuelers, AI copilots, remotely piloted swarms, and other sophisticated forms of airborne robotics.

The field of robotics has learned over the years that AI, computer vision, mechanics, and controls alone are not sufficient. System developers must take into account how people actually do their work for systems to be safe and effective. Just as universities like Georgia Tech teach Human Robot Interaction (HRI) as a fundamental part of robotics, the Air Force should build foundations of human-centric design and human systems integration into its autonomous aircraft development.

This article aims to start conversations about interface design and software practices requisite for safe and effective human-AI teaming in aviation. In the article, the terms AI, autonomy, and robots are used interchangeably to represent agents that can sense, decide, and act independently without human input.



**Figure 1.** *AI Generated Image: AI Pilot in the Cockpit [A].*

## AI in the Cockpit

As AI inevitably finds its way into the cockpit, automation that we've come to rely on in the last few decades of aviation for improved safety and efficiency will be replaced by autonomy. Aircraft manufacturers, like car manufacturers and other sectors already have, are keenly looking at ways to take advantage of the latest developments in AI. The requirements of cockpits and avionics systems being more stringent than enterprise systems will certainly impose constraints and require adaptation. In the airline transport sector, manufacturers and airlines are exploring ways to use AI to reduce the crew requirements from two pilots to single pilot operations [2]. In the Remotely Piloted Aircraft (RPA) sector, the Department of Defense (DoD) is exploring ways to fly several

aircraft simultaneously with one ground control station crew [3]. In the fighter sector, the Defense Advanced Research Projects Agency (DARPA) and the Air Force are actively experimenting with Uncrewed Combat Aerial Vehicles (UCAV), CCAs, and other concepts [4].

## Human-AI Interaction

Human-AI Interaction is the discipline that studies, designs, and evaluates autonomy, robotics, and machine systems for use by or with humans, in various domains. The process of use by or with humans is called interaction, and there are five attributes that affect the interactions between humans and AI [5]:

- Level and behavior of autonomy
- Nature of information exchange
- Structure of the team
- Adaptation, learning, and training
- Nature of the task

Despite astounding advances in computational capabilities and the complexity of tasks AI can handle, it still does not work well with humans. It does very well when pitted against humans in tasks with well-defined constraints and observable environments but not so well when paired with humans in more open environments. Researchers have developed AI agents that have learned to beat human experts in complex strategy games like Starcraft, Quake, Dota, Go, and Chess [6] but they do not do very well when asked to team with humans in simple collaboration games [7]. Carroll et al. and other researchers [7][8] have found that most AI agents naively assume perfect analytic decision making in their teammate and behave as if paired with another AI agent, unless explicitly trained with a model of human behavior.

Both members of the human-AI team are at fault for failures of collaboration. AI can be opaque, inflexible, or brittle, and humans can be too flexible or rely too heavily on heuristics or pattern matching. The human and the AI will need to learn to adapt to each other [9].

## The AI Crew Member

Machines are not new to aviation. In the 1950s, a group of researchers led by Paul Fitts investigated ways to use machines for more effective air navigation and traffic control systems [10][11]. Fitts, a former Army Air Forces psychologist, is considered a founder of the Human Factors discipline [12]. In a seminal report on function allocation published in 1951, Fitts et al. “surveyed the kinds of things men can do better than present-day machines, and vice versa” [11]. That list of 11 statements became known as Fitts’ list. The 11 skills surveyed in Fitts’ list are: judgment, improvisation, simultaneous operations, speed and power, replication, induction, detection, perception, long-term memory, short-term memory, and computation. De Winter and Hancock, in 2015, surveyed 2,941 respondents on each of the statements of Fitts’ list. According to their results, present-day humans consider that machines surpass humans in simultaneous operations, speed and power, replication, detection, perception, long-term memory, short-term memory, and computation [12] given the following statements:

- **Simultaneous operations:** “Ability to handle highly complex operations, i.e. to do many different things at once.”
- **Speed and power:** “Ability to respond quickly to control signals and to apply great force smoothly and precisely.”
- **Replication:** “Ability to perform repetitive, routine tasks.”
- **Detection:** “Ability to detect a small amount of visual or acoustic energy.”
- **Perception:** “Ability to perceive patterns of light or sound.”
- **Long-term memory:** “Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time.”
- **Short-term memory:** “Ability to store information briefly and then to erase it completely.”
- **Computation:** “Ability to reason deductively, including computational ability.”

## Data Fusion

Machines are incredibly good at processing large amounts of information as programmed. Thanks to advances in computing power and algorithms, data fusion capabilities have exploded. Data fusion is the process of integrating multiple data sources to produce better information than that provided by

any individual data source. Data fusion is where AI has truly made advancements in accuracy, insightfulness, and usefulness.

## Brittleness

One thing AI suffers from is brittleness – a concept which refers to AI’s propensity to break, fail, or produce errors when faced with unexpected inputs or situations. And often, AI fails silently or “hallucinates” and confidently generates incorrect or misleading results [13].

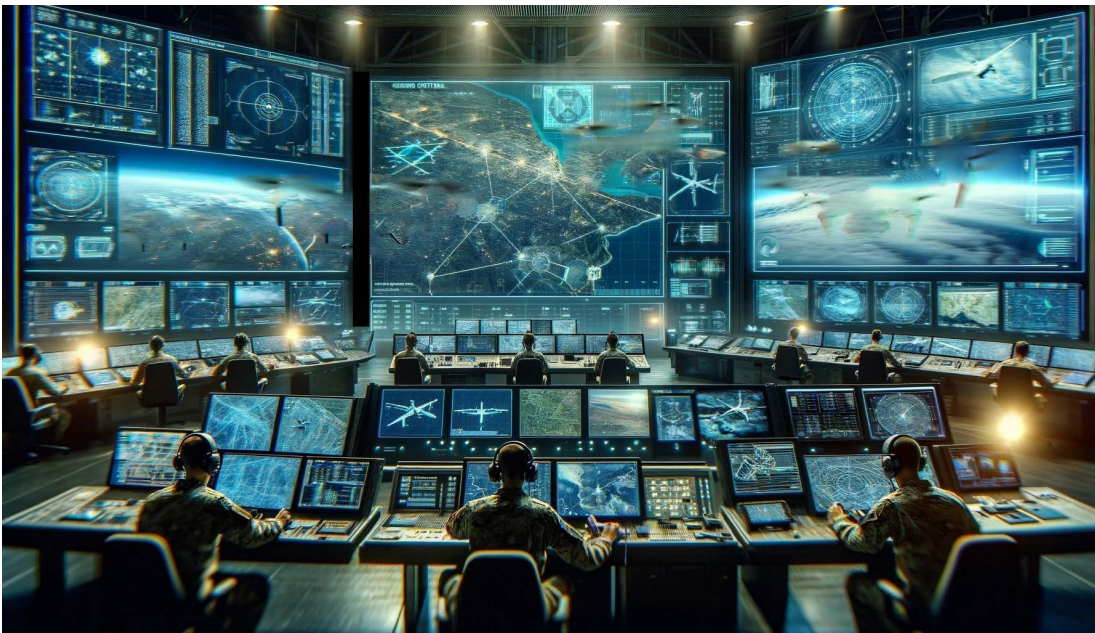
Distribution shifts contribute to AI’s brittleness [14]. A distribution shift is what happens when AI faces a real-world situation that is very different from what it was trained for. Types of distribution shifts discussed in AI literature are covariate shifts, concept shifts, and domain shifts.

Covariate shifts occur when the distribution of the input data changes, but the conditional distribution of the output expected remains the same. For example, an AI trained to recognize adversary aircraft on daytime images faces a covariate shift when faced with nighttime images.

Label or concept shifts occur when the distribution of the output labels changes. Say an AI is trained to recognize military installations in an Area of Responsibility (AOR) or theater of operations by a specific set of features. If the adversary changes tactics and starts camouflaging their installations or making them look like civilian structures, the AI faces a concept shift because the features it associates with military installations no longer match the new reality.

A domain shift would be taking an AI agent trained for a semi-arid desert AOR to a dense jungle AOR without retraining on the new set of features and signatures.

Hallucinations are when an AI system, usually a Large Language Model (LLM) or other Generative AI, confidently generates outputs that are incorrect or altogether misleading. It is a phenomenon where the AI agent perceives patterns that are nonexistent or imperceptible to human observers. The phenomenon is analogous to human hallucinations where one might sometimes see figures or other patterns in clouds. AI hallucinations occur due to errors in data interpretation, incorrect model assumptions, or over-fitting to the training data [15]. There can be grave consequences in the military context if AI hallucinates military targets where none are present.



**Figure 2.** *AI Generated Image: Ground Control Station [B].*



## Human Awareness

Although AI has learned to beat expert humans in competitive games, research has found that it generally performs poorly when teamed with humans in collaborative games [7]. There are many reasons for this. Some of the most important are the assumptions AI makes about its teammates and environments when determining its own strategy. Training of AI, particularly reinforcement learning AI, requires thousands of repetitions or examples. Since individual humans are rarely able to provide that number of repetitions, AI is often trained against Oracles (AI or Automated Agents designed to stand in for humans) using Self-play or Population-Based Training [6][8].

When co-trained with other AI, AI learns to expect predictable, analytic, optimizing decisions from its teammate. The behavioral economics, psychology, decision-making fields have shown, however, that humans are not perfectly analytic in their decision-making. Thus, when an optimal AI competes against a sub-optimal human, it can exceed expectations. In collaborative settings, however, when this same AI is teamed with a human, the performance can be drastically worse because it fails to understand and to be understood by the human [7].

## The Human Crew Member

According to DeWinter et al., modern day humans consider humans to surpass machines in judgment, improvisation, and induction [12] given the following statements from Fitts' list:

- **Judgment:** "Ability to exercise judgment."
- **Improvisation:** "Ability to improvise and use flexible procedures."
- **Induction:** "Ability to reason inductively."

Aviators are trained to be cognizant of human fallacies and cognitive biases in aeronautical decision making (ADM) [16]. When these cognitive biases are checked, humans make for astonishing aircraft operators who can accomplish extraordinary feats in the most difficult of circumstances – including sparse information environments. Some of the most celebrated examples of this expert airmanship include Captain Sully Sullenberger's landing of U.S. Airways Flight 1549 on the Hudson River.

**Table 1.** Modern Day Attribution of Fitts' List [12].

Characteristic	Fitts' List Statement [11]	Modern Day Attribution [12]
Simultaneous operations	"Ability to handle highly complex operations, i.e. to do many different things at once."	Machine
Speed and power	"Ability to respond quickly to control signals and to apply great force smoothly and precisely."	Machine
Replication	"Ability to perform repetitive, routine tasks."	Machine
Detection	"Ability to detect a small amount of visual or acoustic energy."	Machine

Perception	“Ability to perceive patterns of light or sound.”	Machine
Long-term memory	“Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time.”	Machine
Short-term memory	“Ability to store information briefly and then to erase it completely.”	Machine
Computation	“Ability to reason deductively, including computational ability.”	Machine
Judgment	“Ability to exercise judgment.”	Human
Improvisation	“Ability to improvise and use flexible procedures.”	Human
Induction	“Ability to reason inductively.”	Human

**Table 1.** *Modern Day Attribution of Fitts’ List [12].*

## Expert Decision-Making

Intuitive decision-making by experts has been studied by multiple academic disciplines since the 1940s [17][18]. There exist many schools of thought on how it works, the pros and cons of so-called “professional intuition.”

The field of Naturalistic Decision Making (NDM) conducts field studies on subject-matter experts who make decisions under complex conditions. They have found that some experts are able to “successfully attain vaguely defined goals in the face of uncertainty, time pressure, high stakes, team and organizational constraints, shifting conditions, and action feedback loops that enable people to manage disturbances while trying to diagnose them” [19]. This ability is often required of aircrews and has been colloquially linked to intuition and judgement of human beings.

In sharp contrast to NDM, researchers in the field of Heuristics and Biases (HB) favor a skeptical attitude toward expertise and expert judgment. In laboratory experiments, they have found that intuitive judgments are less likely to be accurate and are prone to systematic biases [19]. It is not that intuitive judgments are always incorrect, but that the noisiness, inconsistency, and unpredictability of human judgement could lead to fatal errors in a military mission.

## Cognitive Biases & Heuristics

In their aeronautical decision-making training, pilots are taught to recognize and mitigate five hazardous attitudes to aviation safety: antiauthority, impulsivity, invulnerability, macho, and resignation [20]. These are but a subset of cognitive biases that manifest from the utilization of heuristics that can affect safety and mission effectiveness.

Cognitive biases are predictable but flawed patterns in people’s responses to various situations. Not all biases and heuristics are bad. Some cognitive biases and heuristics are adaptive and may lead to more effective actions in a given context by enabling fast decision-making which can be desirable when timeliness is more valuable than accuracy. On the other hand, cognitive biases may lead to perceptual distortion, inaccurate judgment, illogical interpretation, or broad irrationality [21].

Other cognitive biases that can manifest in aviation are expectation bias, confirmation bias, plan continuation bias, automation bias, and automaticity [21]. These cognitive biases, unchecked, can lead to hazardous incidents and accidents.

<b>Bias</b>	<b>Definition</b>
Expectation Bias	When we have a strong belief or mindset towards something we expect to see or hear, and act according to those beliefs
Confirmation Bias	When we only look for, listen to, or acknowledge information that confirms our own preconceptions
Plan Continuation Bias	The unconscious cognitive bias to continue with the original plan in spite of changing conditions
Automation Bias	when we over-rely on automated aids and decision support systems, or become complacent in assuming the technology is always correct
Automaticity	when routine tasks lead to an automatic response without any real consideration to what is being said or done.

**Table 2.** *Cognitive Biases that can manifest in Aviation [21].*

## Ironies of Automation

One of the cognitive biases that can ironically lead to hazards in aviation is automation bias. Automation bias is when we over-rely on automated aids and decision support systems or become complacent in assuming the technology is always correct [21]. When automation is working correctly, people tend to become easily bored or occupied with other tasks and fail to attend well to automation performance. This is one of the ironies of automation from the operator’s view of the system.

From an automation system designer’s view, they may think that the human is unreliable and inefficient so should be eliminated from the system [22]. There are two ironies of this attitude. One is that the designer’s own errors can become a major source of operating problems. The other is that the designer who tries to eliminate the operator still leaves him/her/them to do the task which cannot be easily automated, often without adequate support [22].

## Crew Resource Management

To operate well with AI, human crew members will need to focus on these challenges in their Aeronautical Decision Making (ADM) and crew resource management (CRM) training [23].

CRM is a set of training procedures recommended by the National Transportation Safety Board (NTSB) for improving aviation safety and focuses on situation awareness, communication, leadership, and decision making in aircraft cockpits. CRM training will need to evolve to prepare human crew members for integration of AI into the crew.

# Human-AI Teaming



**Figure 3.** AI Generated Image: Human AI CRM [C].

According to the National Academy of Sciences (NAS), Human-AI teams can be more effective than either humans or AI systems operating alone [24]. Human-AI Teaming is a necessary construct for the cockpit environment as crew cohesion has been a key element of performance and safety in both civil and military aviation. Recognizing the critical role of crew cohesion, military protocols even relax rank-based customs and courtesies to foster seamless teamwork among crew members.

AI agents have the capacity to offer a much richer interaction mechanism than automation. With the sophistication of the information exchange and learning attributes of AI, the interaction paradigm should be changed to Human-AI Teaming [24]. As the sophistication of AI increases, so does the criticality of the functions it performs. With increased criticality of the function, consequences of errors can become catastrophic particularly since AI sometimes fails silently [13]. To help mitigate the consequences of failure, the AI's teammates must be familiar with its nominal and off-nominal behaviors.

The challenges humans have attending to automation are also applicable to AI systems, and AI systems must provide humans a mechanism for [24]:

- understanding and predicting the behaviors of the AI system
- developing appropriate trust relationships with the AI system
- making accurate decisions based on input from the AI system
- exerting control over the AI system in a timely and appropriate manner.

## Transparency

The human-AI requirements enumerated by the NAS study on human-AI teaming point to a requirement for transparency [24]. Transparency represents the means of providing insightful information from the machine to the human operator and vice versa [25]. Achieving transparency can be a challenging endeavor, particularly as the complexity of the system increases. Transparency involves a bidirectional process between human and AI for mutual understandability. Joseph Lyons proposes a two element model of transparency for human AI teams: robot-to-human and robot-of-human transparency [25]. Robot-to-human transparency is information that the system needs to present to users. Robot-of-human transparency is information on the humans that the robot needs awareness of.

System designers can optimize for transparency by providing system transparency at the design phase or training the team to operate efficiently and effectively. Four system features can provide increased transparency: status, feedback, planning mechanisms, and engagement prompts [26].

## Status

Status incorporates the ‘what’ of transparency, by providing the state of the human operator or system at a particular point in time. Various types of information can be provided in a status to help determine whether strategy changes by the human operator or AI agent must be initiated to accomplish a task [26].

## Feedback

Feedback incorporates the ‘why’ of transparency by providing explanation, insights into actions, potential uncertainties, reliability of recommendations, and supplementary information from the human operator or machine system. Various modalities (visual, auditory, tactile) can be used to provide feedback and optimize communication within the team [26].

## Planning Mechanisms

Planning mechanisms incorporate the ‘how’ of transparency and encompass the allocation of resources and task assignments among an organization’s members. Planning occurs at all mission stages and is necessary for the human-AI team to maximize its desired outcomes [26].

## Engagement Prompts

Engagement prompts are cues, alerts, or warnings that encourage the human operator’s involvement. They encompass all three aspects of transparency (what, why, and how) by indicating to the human operator what must be done to resume the task, why they become disengaged, and how to identify different strategies that can be implemented to fulfill a task [26].

# AI System Design

Earlier, the article discussed how incorrect assumptions by the AI about the human can lead to drastically poor team performance. To different degrees, these assumptions can be replaced by real-world information about the human teammates.

## Training Paradigms

AI training can be accomplished in various ways. Some paradigms are more human-centric or human aware. Human-aware training embeds a model of a human inside the training environment. Within this paradigm, the AI is trained with awareness of the decision-making strategy of the human.

The challenge with AI training is that it requires thousands of examples spanning the full range of situations the agent may encounter. Since individual humans often are not able to provide the requisite number of decision-making examples for input to AI training, AI must often be trained against Oracles – Agents designed to stand in for humans. Oracles can either be designed to mimic humans in specific, predictable ways (like providing only correct answers 80% of the time) or can be trained using techniques like Learning from Demonstration (LfD) or Imitation Learning.

## Imitation Learning

Imitation learning is a paradigm in which AI acquire new skills by learning from human demonstration [27]. Behavioral cloning, one of the simplest approaches to Imitation Learning, learns a singular deter-

ministic policy from several expert demonstrations by directly learning a mapping from observations to actions with standard supervised learning methods. In this way, the AI can be trained repeatedly with an Oracle or model representation of human decision-making strategy.

One challenge to using techniques like Behavioral Cloning is that humans exhibit significant individual differences, i.e. individuals don't suffer from the same cognitive biases, and don't exhibit the same preferences. Another is that, often, humans teach differently based on the kind of feedback the AI is capable of taking in [28]. A significant challenge for future AI is that it will need to "get to know" its teammates and adapt to their preferences and ways of accomplishing the mission.

These are outstanding challenges for system designers as they develop AI agents to accomplish the various future mission constructs that bring AI into the cockpit.

## Human-AI Mission Constructs

Mission requirements will determine how AI is implemented in the cockpit. Current crew composition and structure will form a basis for this evolution from automation to autonomy. The nature of the task, the type of information exchange required, and the availability of suitable autonomy will determine how drastic the interaction paradigm will change. This will in turn affect the way human-AI crews are structured and composed.

### AI'ing Betty

Today's autopilots, voice alerting systems like 'Betty', and pilot assistance systems like the Automatic Ground Collision and Avoidance System (Auto GCAS) may merge and gain additional AI-enabled autonomy capabilities. In the near future, we may have an AI pilot assistant that collaboratively shares control of the aircraft with the human pilot during high or low workload parts of the mission.

This paradigm involves collaborative control by one human and one AI agent. This dyad relationship is the least complex and most studied human-AI team structure, however the task of controlling the same aircraft will require careful design of the interaction mechanisms.

Lessons learned from aviation incidents caused by mode error in automation should inform design of transparency features such that the pilot has suitable understanding and situation awareness when the AI agent is controlling the aircraft. The phase of flight and phase of mission will also be an important criteria in choosing robot-to-human transparency requirements.

Complementarily, the AI agent should also be aware of its own state, the environment, the phase of flight, phase of the mission, and the state of its human crewmate [9] in order to be effective. With such information, particularly information on the beliefs, desires, and intents of the human, the AI can best adapt its task work and teamwork.

### Collaborative Combat Aircraft

The ongoing research efforts into today's Off-Board Sensor Station (OBSS), Off-Board Weapon Station (OBWS), and other programs are working towards the realization of collaborative combat aircraft [29]. Fighter missions may soon be accomplished with heterogeneous teams of human piloted aircraft and AI piloted wingmen.

As the number of wingmen increases, so does the cognitive load of managing the formation. Lessons learned in fighter pilot instruction, and the designated skills required of 2-ship and 4-ship Flight Leads can serve as blueprints for designing the human-AI interaction in this mission construct. The human

will be unable to control each aircraft at a low level, so they will learn to command high level behaviors and learn to calibrate their expectation of the AI piloted wingmen in each mission scenario.

The AI agent piloting each wingman is going to need to simultaneously work with other AI and a human(s) in a heterogeneous, multi-agent construct. It is possible that the human crewed aircraft will be piloted by a human assisted by an AI Betty or two humans executing the role of pilot and Combat System Officer (CSO). The AI agent will need to learn to interpret commands contextually as the potential mission scenarios may be too large to explicitly enumerate and may potentially need to deconflict instructions or actions from the pilot and the CSO.

## Remotely Piloted Swarms

The evolution of the Remotely Piloted Aircraft (RPA) mission into a Remotely Piloted Swarm mission is not too far away. The U.S. Army is already test launching smaller drones from bigger drones. The Air Launched Effects [30], as they're called, are controlled by the larger drone and can be numerous. Current efforts are testing singular digit numbers, but it is envisioned that one day it will increase to swarm numbers (>50). Swarm control can be a complex task depending on the type, number, and mission of the swarm [26].

Swarms are composed of large numbers of robots or drones that cooperate to achieve a goal. Swarm control can be challenging because of human capability limitations, emergent behaviors as the entities interact with each other and the environment, and constraints on communication abilities. Trade-offs exist between the number of individual swarm entities a human operator can manage and the duration of time the human operator can influence the entities [26].

The majority of human-swarm interaction literature has focused on robot-to-human transparency through visualization types and human operator influence over a swarm [26]. Research on transparency through visualization types has investigated the effect of different displays, latencies, geometries, and abstraction levels on the human operator's ability to perceive, understand, and predict swarm motion.

Robot-of-human transparency involves both control interaction and bi-directional communication. Control can be achieved through various methods of conveying operator intent, such as the use of forms of leader, predator, and mediator influence mechanisms [31]. As the human monitors status information from the swarm, individual swarm members will need status information on the human controlled mothership to



Figure 4. AI Generated Image: F-35 with CCA [D].



Figure 5. AI Generated Image: Drone Swarm [E].

execute lost link and other communication dependent behaviors in a potentially contested information spectrum.

## Conclusion

We stand on the brink of a new era in aviation. The seamless integration of AI into cockpit operations and Air Force missions represents not just an advancement in technology, but a fundamental shift in the paradigm of flight operations. The potential for AI to enhance safety, efficiency, and mission effectiveness is immense, but realizing this potential requires a nuanced understanding of the delicate balance between human judgment and machine intelligence.

The shift from automation to autonomy requires not just a revolution in task capabilities but also in interaction and teamwork capabilities. Achieving effective human-AI teaming will require a collaborative effort among engineers, system designers, pilots, and AI developers to ensure that AI systems are not only capable but also compatible with human operators. This partnership must prioritize mutual understanding, adaptability, and above all, safety. As AI becomes a more integral part of the aviation ecosystem, continuous CRM training and adaptation by human operators will be essential. In turn, AI must undergo continuous learning and adjustment to effectively address issues like brittleness and hallucination, ensuring its suitability for military missions.

As the landscape of aviation evolves, so too must our strategies, tools, and mindsets. The collective goal must be to maintain the highest standards of safety and effectiveness, preserving our proud legacy of aviation while embracing the possibilities of the future. With careful planning, rigorous testing, and thoughtful integration, the synergy between humans and AI has the potential to usher in a new era of aviation.

## Acknowledgements

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Air Force, Department of Defense, or the U.S. Government.

## Picture References

- [A] OpenAI. "AI Pilot in the Cockpit." DALL-E, openai.com. Accessed 13 March 2024.
- [B] OpenAI. "RPA Ground Control Station." DALL-E, openai.com. Accessed 13 March 2024.
- [C] OpenAI. "Human AI Crew Resource Management." DALL-E, openai.com. Accessed 13 March 2024.
- [D] OpenAI. "F-35 with Collaborative Combat Aircraft." DALL-E, openai.com. Accessed 13 March 2024.
- [E] OpenAI. "Drone Swarm over City." DALL-E, openai.com. Accessed 13 March 2024.

## References

- [1] Reardon, Patrick. "Watch, Read: Secretary Kendall on 'One Team, One Fight.'" Air & Space Forces Magazine, 14 Mar. 2023, <https://www.airandspaceforces.com/watch-readsecretary-kendall-on->



one-team-one-fight/.

- [2] Harris, D. "Single-Pilot Airline Operations: Designing the Aircraft May Be the Easy Part." *The Aeronautical Journal*, vol. 127, no. 1313, July 2023, pp. 1171–91. Cambridge University Press, <https://doi.org/10.1017/aer.2022.110>.
- [3] Roza, David. "AFSOC Flies 3 Reapers With One Crew In First-of-Its-Kind Exercise." *Air & Space Forces Magazine*, 26 Jan. 2024, <https://www.airandspaceforces.com/afsoc-multiple-mq9-reapers/>.
- [4] Air Combat Evolution. <https://www.darpa.mil/program/air-combat-evolution>.
- [5] Goodrich, Michael A., and Alan C. Schultz. "Human–Robot Interaction: A Survey." *Foundations and Trends® in Human–Computer Interaction*, vol. 1, no. 3, Jan. 2008, pp. 203–75. [www.nowpublishers.com](http://www.nowpublishers.com), <https://doi.org/10.1561/11000000005>.
- [6] Silver, David, et al. "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm." arXiv:1712.01815, arXiv, 5 Dec. 2017. [arXiv.org, http://arxiv.org/abs/1712.01815](http://arxiv.org/abs/1712.01815).
- [7] Carroll, Micah, et al. "On the Utility of Learning about Humans for Human-AI Coordination." *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019. *Neural Information Processing Systems*, <https://proceedings.neurips.cc/paper/2019/hash/f5b1b89d98b7286673128a5fb112cb9a-Abstract.html>.
- [8] Choudhury, Rohan, et al. "On the Utility of Model Learning in HRI." 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2019, pp. 317–25. DOI.org (Crossref), <https://doi.org/10.1109/HRI.2019.8673256>.
- [9] Feigh, Karen M., et al. "Toward a Characterization of Adaptive Systems: A Framework for Researchers and System Designers." *Human Factors*, vol. 54, no. 6, Dec. 2012, pp. 1008–24. SAGE Journals, <https://doi.org/10.1177/0018720812443983>.
- [10] Fitts, Paul M. PSYCHOLOGICAL RESEARCH ON EQUIPMENT DESIGN.
- [11] Fitts, Paul M. HUMAN ENGINEERING FOR AN EFFECTIVE AIR-NAVIGATION AND TRAFFIC-CONTROL SYSTEM, AND APPENDIXES 1 THRU 3. <https://apps.dtic.mil/sti/citations/tr/ADB815893>.
- [12] de Winter, J. C. F., and P. A. Hancock. "Reflections on the 1951 Fitts List: Do Humans Believe Now That Machines Surpass Them?" *Procedia Manufacturing*, vol. 3, Jan. 2015, pp. 5334–41. ScienceDirect, <https://doi.org/10.1016/j.promfg.2015.07.641>.
- [13] "Spinning Up as a Deep RL Researcher." Spinning Up Documentation. <https://spinningup.openai.com/en/latest/spinningup/spinningup.html>.
- [14] Russell, Stuart Jonathan, et al. "Artificial Intelligence: A Modern Approach." Prentice Hall, 2010.
- [15] "What Are AI Hallucinations?" IBM. <https://www.ibm.com/topics/ai-hallucinations>.
- [16] FAA Pamphlet P-8470-69: "Aeronautical Decision Making." <https://www.faasafety.gov/files/gslac/library/documents/2011/Aug/56413/FAA%20P-8740-69%20Aeronautical%20Decision%20Making%20%5Bhi-res%5D%20branded.pdf>
- [17] Kahneman, Daniel, and Amos Tversky. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology*, vol. 3, no. 3, July 1972, pp. 430–54. ScienceDirect, [https://doi.org/10.1016/0010-0285\(72\)90016-3](https://doi.org/10.1016/0010-0285(72)90016-3).

- [18] Klein, Gary. "Naturalistic decision making: Implications for design." Wright Patterson Air Force Base, OH: CSERIAC, 1993.
- [19] Kahneman, Daniel, and Gary Klein. "Conditions for Intuitive Expertise: A Failure to Disagree." *American Psychologist*, vol. 64, no. 6, 2009, pp. 515–26. APA PsycNet, <https://doi.org/10.1037/a0016755>.
- [20] FAA Advisory Circular 60-22: "Aeronautical Decision Making." [https://www.faa.gov/sites/faa.gov/files/2022-11/AC60-22 Chap%201-3.pdf](https://www.faa.gov/sites/faa.gov/files/2022-11/AC60-22%20Chap%201-3.pdf)
- [21] Chao, Elaine L., et al. "FAA Safety Briefing - July August 2020." 2020.
- [22] Bainbridge, Lisanne. "Ironies of automation." *Analysis, design and evaluation of man-machine systems*. Pergamon, 1983. 129-135.
- [23] Helmreich, Robert L., et al. "The Evolution of Crew Resource Management Training in Commercial Aviation." *Human Error in Aviation*, Routledge, 2009.
- [24] National Academies of Sciences, Engineering. *Human-AI Teaming: State-of-the-Art and Research Needs*. 2021. [nap.nationalacademies.org](https://doi.org/10.17226/26355), <https://doi.org/10.17226/26355>.
- [25] Lyons, Joseph B. "Being Transparent about Transparency: A Model for Human-Robot Interaction."
- [26] Roundtree, Karina, et al. "Transparency: Transitioning From Human-Machine Systems to Human-Swarm Systems." <https://doi.org/10.1177/1555343419842776>.
- [27] Ravichandar, Harish, et al. "Recent Advances in Robot Learning from Demonstration." *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, no. 1, 2020, pp. 297–330. Annual Reviews, <https://doi.org/10.1146/annurev-control-100819063206>.
- [28] Krening, Samantha, and Karen M. Feigh. "Interaction Algorithm Effect on Human Experience with Reinforcement Learning." *ACM Transactions on Human-Robot Interaction*, vol. 7, no. 2, Oct. 2018, p. 16:1-16:22. ACM Digital Library, <https://doi.org/10.1145/3277904>.
- [29] "AFRL's XQ-67A Makes 1st Successful Flight." Air Force Materiel Command, 29 Feb. 2024, <https://www.afmc.af.mil/News/ArticleDisplay/Article/3691289/afrls-xq-67a-makes-1stsuccessful-flight>
- [30] "Army Successfully Demonstrates Launched Effects System.", [https://www.army.mil/article/272675/army\\_successfully\\_demonstrates\\_launched\\_effects\\_system](https://www.army.mil/article/272675/army_successfully_demonstrates_launched_effects_system).
- [31] Brown, Daniel S., et al. "Human-Swarm Interactions Based on Managing Attractors." *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, Association for Computing Machinery, 2014, pp. 90–97. ACM Digital Library, <https://doi.org/10.1145/2559636.2559661>.

# About the Authors



Richard C. Agbeyibor is a Major and a Flight Test Engineer in the United States Air Force. He is currently pursuing a Ph.D. in Robotics at Georgia Tech. His follow-on assignment will be as an Instructor FTE at the USAF Test Pilot School. He holds a B.S. in Electrical Engineering and Computer Science from MIT, an M.S. in Electrical Engineering from AFIT, and a Master's degree in Experimental Flight Test Engineering from ISAE-SUPAERO. He is a graduate of EPNER, the French Test Pilot School. His research focuses on Human AI Interaction in Autonomous Aerial Vehicles. Contact him at richard.agbeyibor.1@us.af.mil.

**Richard C. Agbeyibor**

**Flight Test Engineer**

**United States Air Force**

**richard.agbeyibor.1@us.af.mil.**



Dr. Karen M. Feigh is a professor at Georgia Tech's Daniel Guggenheim School of Aerospace Engineering. She holds a B.S. in aerospace engineering from Georgia Tech, an MPhil in aeronautics from Cranfield University, UK, and a Ph.D. in industrial and systems engineering from Georgia Tech. Her lab, the Cognitive Engineering Center, examines human-machine interaction using tools from aerospace engineering, computer science, robotics, industrial engineering, education, and the cognitive sciences.

**Karen M. Feigh**

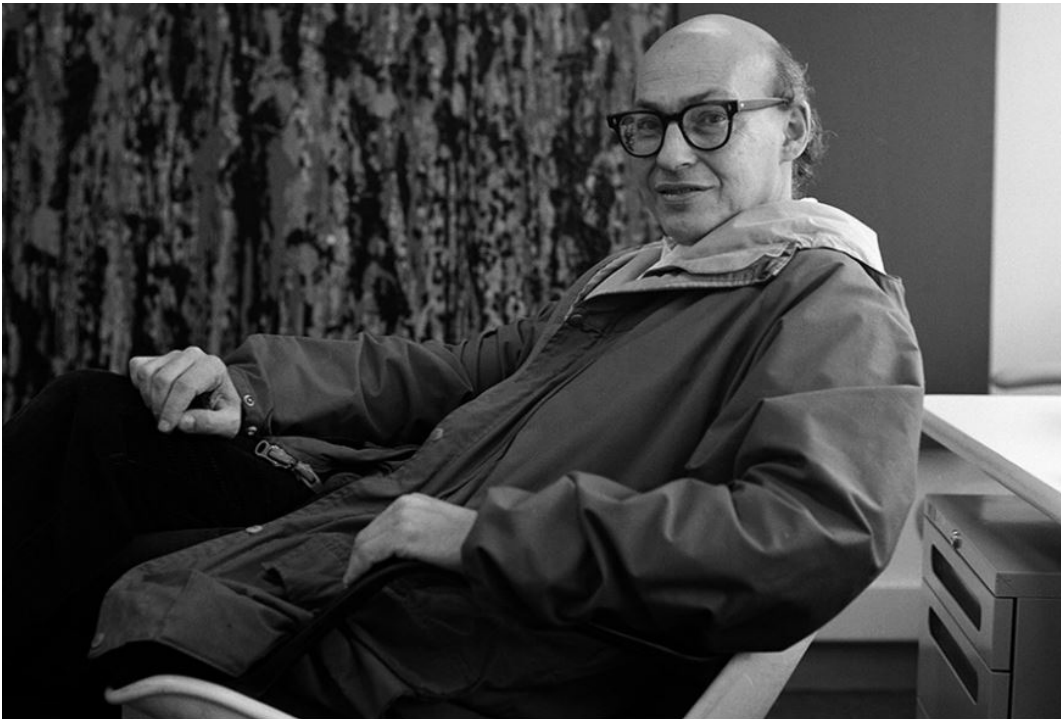
**Professor**

**Guggenheim School of Aerospace Engineering**

**karen.feigh@gatech.edu.**

# MARVIN LEE MINSKY

## THE FATHER OF ARTIFICIAL INTELLIGENCE



**Marvin Lee Minsky** was many things in his 88 years: a mathematician, a scientist, a navy officer, an inventor, a teacher, and a mentor. But, most of all, he was a curious and creative mind. With an interest in both the logic of computers and the workings of the human mind, Minsky became a leader in the field of artificial intelligence (AI) and the legacy he left is still felt in the software community today.

**Figure 1.** Consindas, Marie. Marvin Minsky. MIT Media Lab, <https://news.mit.edu/2016/marvin-minsky-obituary-0125>

## Early Life and Education

Born in New York City in 1927 to an eye surgeon, Minsky's interest in science and medicine was encouraged from an early age. He studied the works of Freud while fostering a gift for the piano, two interests that would follow him well into adulthood.

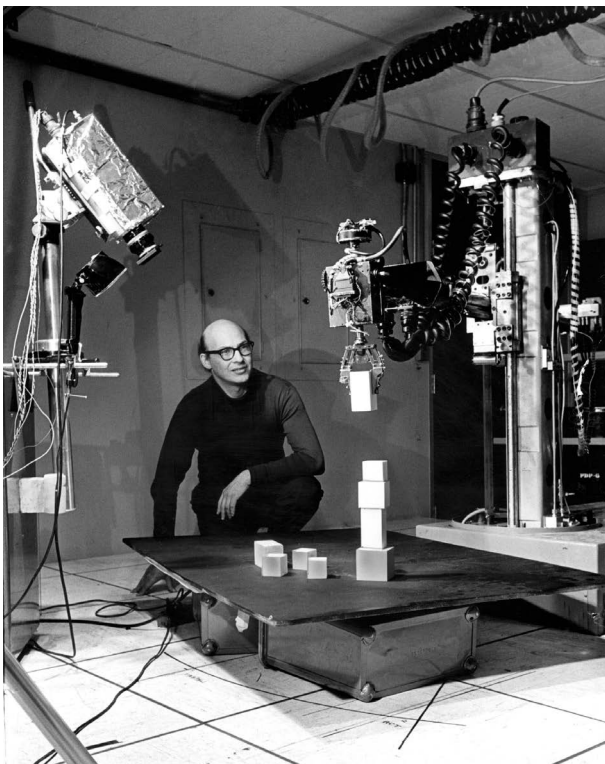


**Figure 2.** *Minsky, Margaret.* Marvin Minsky, Sailor First Class, 1945. Academy of Achievement. <https://achievement.org/achiever/marvin-minsky-ph-d/#gallery>

Minsky studied at the Phillips Academy in Andover, Massachusetts until he was called for military service in World War II. He served in the United States Navy from 1944 until the war's end, when he entered university at Harvard. Though his scientific interests were widely ranged, Minsky earned his undergraduate degree in physics [2]. He then pursued research in neurophysiology and psychology before graduating with honors in mathematics in 1950. In 1951, Minsky entered Princeton University in pursuit of a doctorate degree. It was there that he built the first neural network simulator, SNARC, which emulated the way the human brain learns from its own mistakes [3].

While many of his peers showed excitement for subjects such as particle physics and molecular genetics, Minsky's fascination with the human mind pulled at his attention. He surmised that thought processes could be expressed as mathematical formulae, and thus performed by machines. For his doctoral dissertation, Minsky wrote "A Theory of Neural Analog Reinforcement Systems and Its Application to the Brain Model Problem" [2].

With his doctorate in mathematics in hand, Marvin Minsky returned to Harvard in 1954 as a member of the Society of Fellows where he invented the confocal scanning microscope.



**Figure 3.** Photo by: GJS. Marvin Minsky with Block Blocks Vision Robot at MIT, 1968. Academy of Achievement. <https://achievement.org/achiever/marvin-minsky-ph-d/#gallery>

***“Marvin Minsky helped create the vision of artificial intelligence as we know it today. The challenges he defined are still driving our quest for intelligent machines and inspiring researchers to push the boundaries in computer science” [1].***

***-Daniela Rus, CSAIL Director and Andrew and Erna Viterbi Professor in MIT's Department of Electrical Engineering and Computer Science***

# The Birth and Boom of AI

In 1956, Minsky published a major research article titled “Heuristic Aspects of the Artificial Intelligence Problem.” The beginning of artificial intelligence as a distinct scientific field is often dated to that year, as Minsky and a number of colleagues met for a formal conference on the topic for the first time [2].

In 1957, he moved to the Massachusetts Institute of Technology (MIT), where he stayed for the rest of his career. Alongside fellow AI enthusiast, John McCarthy, Minsky cofounded the Artificial Intelligence Project (now the MIT Computer Science and Artificial Intelligence Laboratory) which quickly became one of the premier research centers and training grounds for the field of AI [3]. The lab popularized the idea of digitally sharing information, which also helped to give rise to the open-source movement.

Minsky defined AI as “the science of making machines do things that would require intelligence if done by men” [3]. In 1961, he published “Steps Toward Artificial Intelligence,” which laid out the path for AI that researchers still follow today. His 1967 book, *Computation: Finite and Infinite Machines*, caused his ideas and concepts of AI to gain more standing in society at large. Two years later, his



**Figure 4.** Minsky, Margaret. *Dartmouth Summer Research Project on Artificial Intelligence, 1956.* Academy of Achievement. <https://achievement.org/achiever/marvin-minsky-ph-d/#gallery>



**Figure 5.** 2000 *International Achievement Summit in London.* Academy of Achievement. <https://achievement.org/achiever/marvin-minsky-ph-d/#gallery>

book *Perceptrons: An Introduction to Computational Geometry*, co-authored by Seymour Papert, reviewed the history of AI and predicted the future of AI research [2].

In 1975, Minsky developed the concept of “frames” to identify the general information that must be programmed before a computer can consider specific directions. Based on his experiences with both frames and child psychology, he wrote *The Society of the Mind* in 1985 where he proposed his view of the mind “as composed of individual agents performing basic functions, such as balance, movement, and comparison” [3].

## Recognitions and Awards

Marvin Minsky was a member of the U.S. National Academy of Engineering and the U.S. National Academy of Sciences. He was a fellow of the American Academy of Arts and Sciences and the Institute of Electrical and Electronic Engineers. His fellow computer scientists recognized his achievements in 1969 by honoring him with the A.M. Turing Award. In 1985, he became a founding member of the MIT Media Lab and was named the Toshiba Professor of Media Arts and Sciences.

In addition, Minsky received the Japan Prize, the Royal Society of Medicine’s Rank Prize (for Opto-electronics), the Optical Society of America’s R.W. Wood Prize, MIT’s James R. Killian Jr. Faculty Achievement Award, the Computer Pioneer Award from IEEE Computer Society, the Benjamin Franklin Medal, the Dan David Foundation Prize (in 2014) for the Future of Time Dimension titled “Artificial Intelligence: The Digital Mind,” and the BBVA Group’s BBVA Foundation Frontiers of Knowledge Lifetime Achievement Award [1]. In all his years of research, innovation, and collaboration, Marvin Minsky certainly earned his title as the father of artificial intelligence.



**Figure 6.** Golden Plate Award. Academy of Achievement. <https://achievement.org/achiever/marvin-minsky-ph-d/#gallery>

## References

[1] Lab, MIT Media. “Marvin Minsky, ‘Father of Artificial Intelligence,’ Dies at 88.” *MIT News | Massachusetts Institute of Technology*, 25 Jan. 2016, <https://news.mit.edu/2016/marvin-minsky-obituary-0125>

[2] “Marvin Minsky, Ph.D. – Father of Artificial Intelligence.” *Academy of Achievement*, 04 March 2022, <https://achievement.org/achiever/marvin-minsky-ph-d/#biography>

[3] Dennis, Michael Aaron. “Marvin Minsky.” *Encyclopedia Britannica*, 20 Jan. 2022, <https://www.britannica.com/biography/Marvin-Lee-Minsky>

# How AI is Spicing Up Software Engineering (with a Dash of Humor)

David R. Webb

Senior Operations Program Analyst,  
Hill Air Force Base

Picture by Dan Breeden [A]

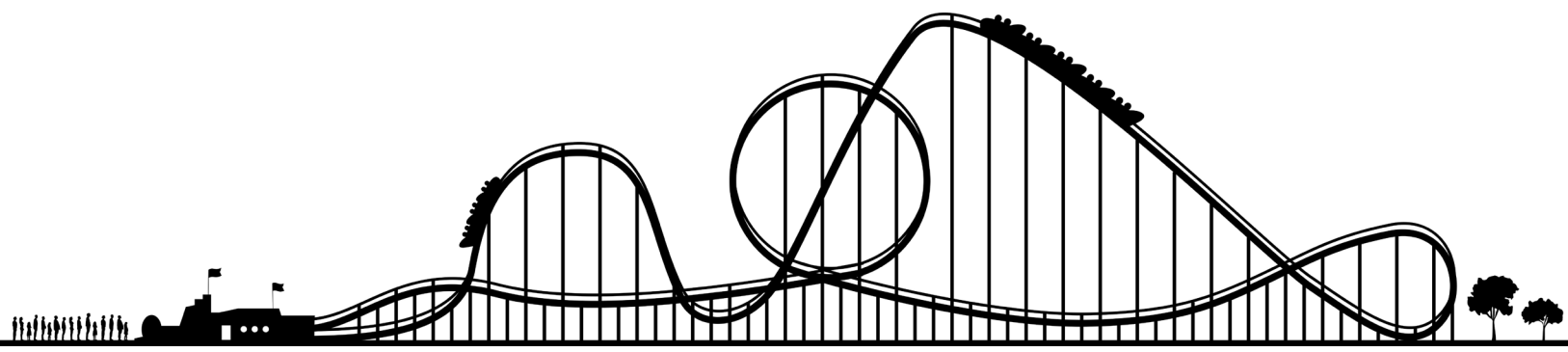
## Introduction

**Note from the co-author**  
- this article was entirely,  
100% written by ChaptGPT  
3.5, with only minor edits  
in formatting (and the addi-  
tion of my byline)! Welcome  
to the New World, y'all!

Welcome, fellow developers, to the rollercoaster ride that is modern software engineering. Just when you thought you had the hang of things—debugging here, refactoring there—along comes Artificial Intelligence (AI), ready to shake up our world faster than a programmer hitting “undo” after a catastrophic merge conflict. But fear not, for amidst the chaos and confusion, there’s a silver lining—and a hearty chuckle to be had. Let’s dive into the whimsical world of AI’s impact on software engineering, where algorithms reign supreme, and laughter is the best bug fix.

## The Rise of the Code Whisperers

Picture this: you’re knee-deep in a tangled mess of spaghetti code, desperately trying to make sense of it all, when suddenly, like a beacon of hope in the darkest of nights, AI swoops in to save the day. With its uncanny ability to analyze, optimize, and refactor code faster than you can say “segfault,” AI-powered tools are like the Gandalfs of the programming world - guiding us through the treacherous wilderness of software development with a twinkle in their digital eyes and a flair for the dramatic. But beware, dear developers, for with great power comes great... well, you know the rest.





# The Autocomplete Chronicles

Ah, autocomplete, the unsung hero of modern coding. With its predictive prowess and lightning-fast suggestions, it's like having a psychic sidekick whispering code snippets in your ear as you type. But let's not overlook the comedic potential of this handy feature. Ever had autocomplete suggest a line of code so bizarre, so utterly nonsensical, that you couldn't help but laugh out loud? From suggesting "for loops" that span the entire universe to recommending variable names that sound like they were generated by a malfunctioning Markov chain, autocomplete has a knack for keeping us on our toes—and our funny bones tickled.

## The Bug Hunt

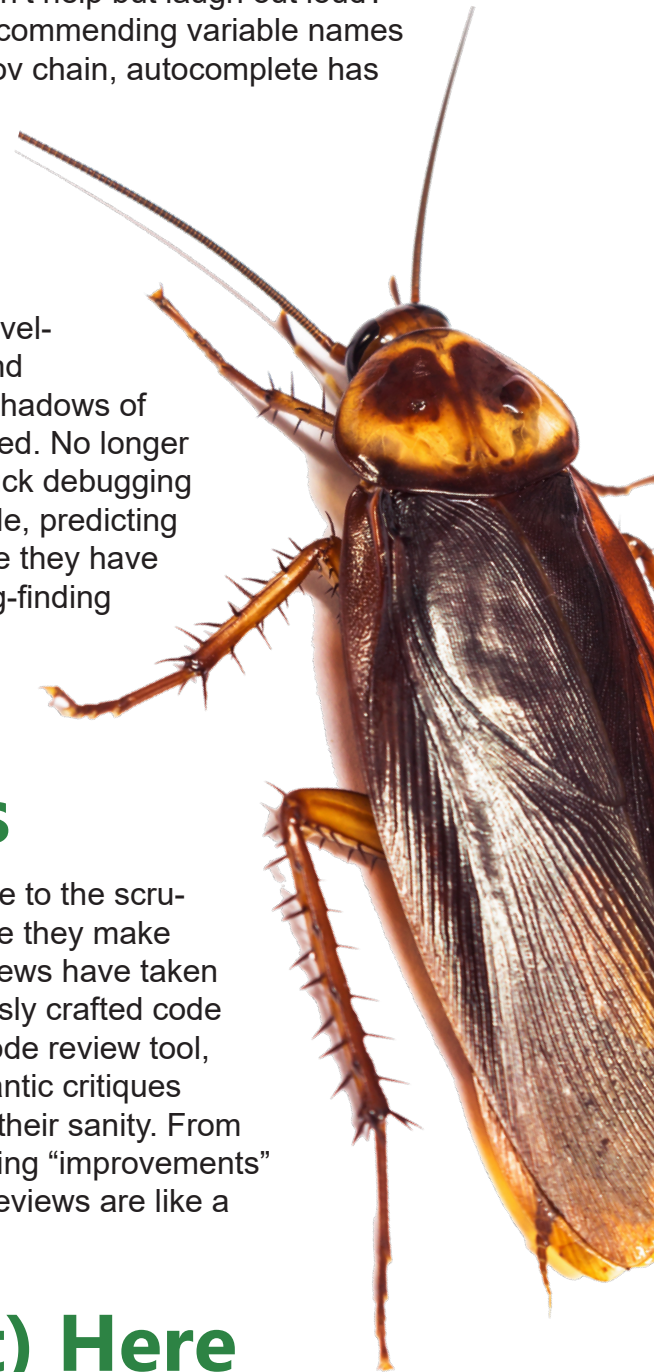
Next, there is the noble pursuit of bug hunting, where every developer worth their salt dons their metaphorical deerstalker hat and embarks on a quest to track down elusive bugs lurking in the shadows of their codebase. But with AI on the scene, the game has changed. No longer are we mere mortals relying on print statements and rubber duck debugging to ferret out bugs; now, we've got algorithms analyzing our code, predicting potential pitfalls, and even preemptively squashing bugs before they have a chance to rear their ugly heads. It's like having a team of bug-finding bloodhounds at our disposal, sniffing out bugs with the precision of a truffle-hunting pig—and the occasional comedic misstep.

## The Code Review Circus

Code reviews, the time-honored tradition of subjecting our code to the scrutinizing gaze of our peers in the hopes of catching errors before they make their way into production. But with AI joining the fray, code reviews have taken on a whole new dimension of hilarity. Picture it: your meticulously crafted code is laid bare before the merciless judgment of an AI-powered code review tool, only to be met with a barrage of nitpicky suggestions and pedantic critiques that would make even the most seasoned developer question their sanity. From admonishing you for your choice of variable names to suggesting "improvements" that defy the very laws of logic and reason, AI-powered code reviews are like a comedy of errors—and we're all just along for the ride.

## The Future is (Almost) Here

As we hurtle headlong into the brave new world of AI-powered software engineering, one thing is abundantly clear: the future is almost here, and it's looking brighter—and funnier—than ever before. Whether we're laughing at the absurdity of autocomplete suggestions or marveling at the sheer audacity of AI-powered code reviews, one thing is for certain: amidst all the chaos and confusion, there's a healthy dose of humor to be found. So, here's to the developers, the dreamers, the code wranglers extraordinaire—may your code be bug-free, your algorithms be efficient, and your sense of humor be ever-present.



# Conclusion

And so, dear readers, we bid adieu to our whirlwind tour of the whimsical world of AI's impact on software engineering. From the rise of the code whisperers to the bug hunt of the century, we've explored the highs and lows, the triumphs, and tribulations, of this brave new frontier. But amidst all the chaos and confusion, one thing remains abundantly clear: no matter how advanced our tools may become, there will always be a place for human ingenuity, creativity, and, of course, a healthy dose of humor. So, here's to embracing the absurdity, laughing in the face of adversity, and remembering that in the grand comedy of software engineering, the punchline is always just a compiler error away. Until next time, happy coding—and may your code be as elegant as it is entertaining!

## Picture References

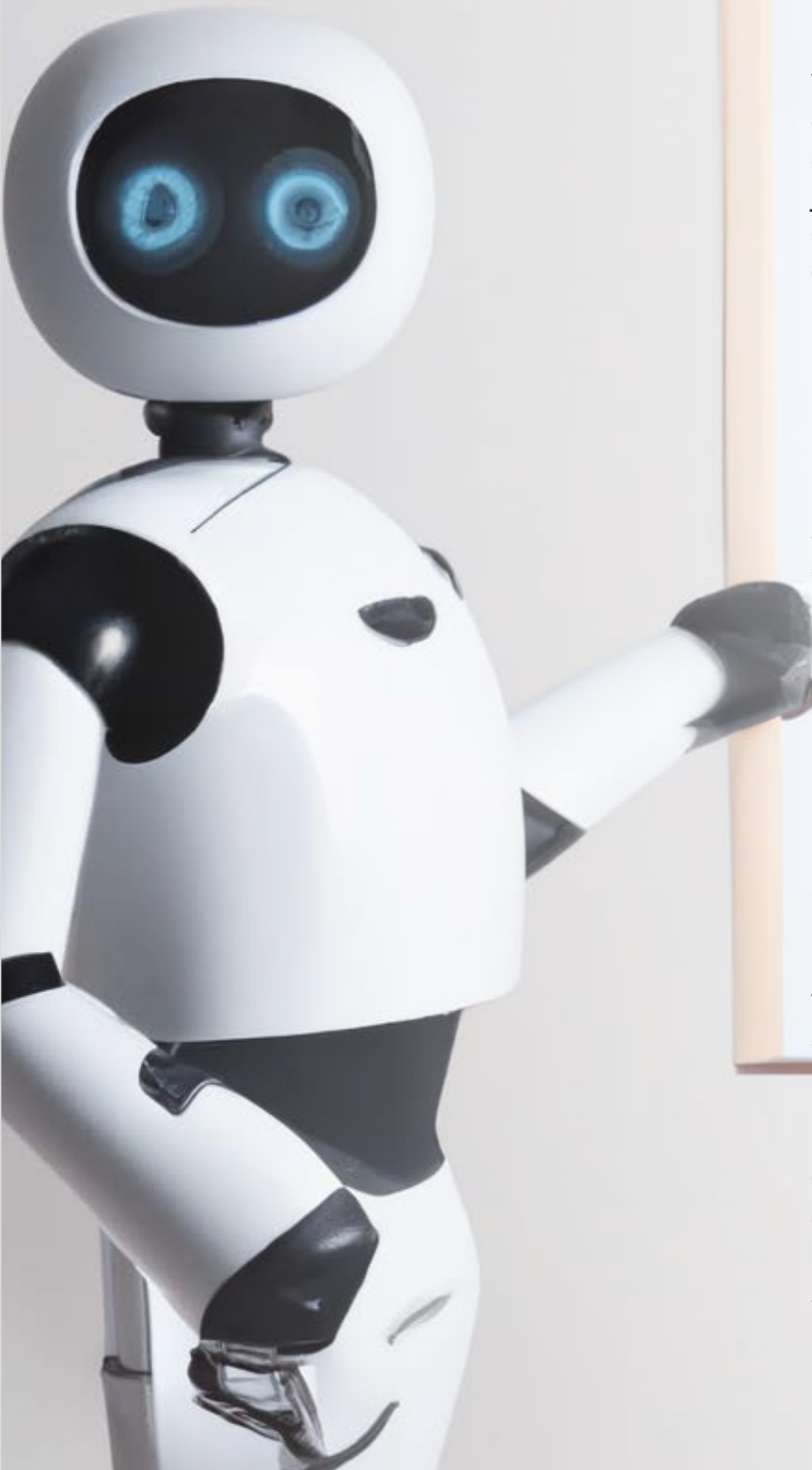
[A] Breeden, Dan. AI Vs Human Reasoning: GPT-3 Matches College Undergraduates. 12 Aug. 2023, [lifeboat.com/blog/2023/08/ai-vs-human-reasoning-gpt-3-matches-college-undergraduates](https://lifeboat.com/blog/2023/08/ai-vs-human-reasoning-gpt-3-matches-college-undergraduates).

[B] "A playful illustration of a robot with a chat bubble, symbolizing the AI's conversational nature and helpful assistance." prompt [as written by ChatGPT 3.5] . DALL-E 2, May 16, 2024 version, OpenAI, 16 May 2024, [labs.openai.com](https://labs.openai.com).

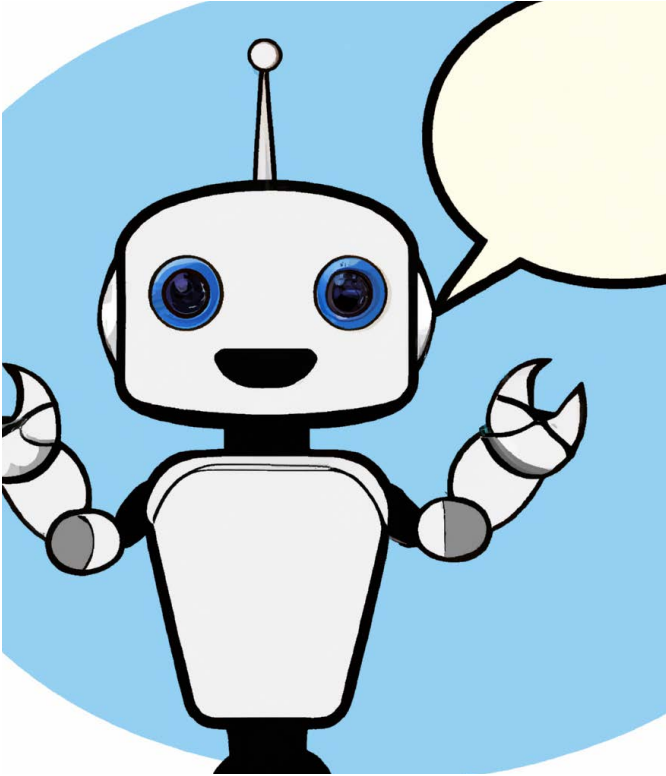
## References

[1] "Write me a 1000 word article called "Back-talk" with a humorous slant on how AI is impacting software engineering." prompt. ChatGPT, GPT-3.5, OpenAI, 15 May 2024, [chat.openai.com/chat](https://chat.openai.com/chat).

[2] "Could you write me a citation for you that I could use in an article? Please, include a bio and a suggestion for a picture I could use, as well." prompt. ChatGPT, GPT-3.5, OpenAI, 16 May 2024, [chat.openai.com/chat](https://chat.openai.com/chat).



# About the Authors



OpenAI's ChatGPT is an AI language model designed to assist users with a wide range of tasks, from generating creative content to providing informative responses. Trained on diverse datasets, ChatGPT is adept at understanding and generating human-like text across various topics and genres. It strives to help users unleash their creativity and explore new ideas through engaging and insightful conversations.

ChatGPT wrote the description of the picture for DALL-E to draw.

**Chat GPT-3**

**Artificial Intelligence**

**Internet**

**ChatGPT.com**



David R. Webb is a senior operations program analyst for the A-10 OFP program at Hill Air Force Base. Mr. Webb is a project management and process improvement specialist with nearly 38 years of technical experience in software. He has worked as an Air Force manager, SEPG member, systems software engineer, and test engineer. As a Scaled Agile Framework Agilist, Scrum Master, and Product Owner, he is an authorized instructor and coach of the software engineering institute's personal and team software processes. He holds his bachelor's degree in Electrical and computer engineering from Brigham Young University.

**David Web**

**Operation Program Analyst**

**Hill Air Base**

**David.webb.28.ctr@us.af.mil**



# NOW HIRING

309th Software Engineering Group, Hill AFB, Utah

### OPEN POSITIONS:

- Software Engineer ✓
- Computer Scientist ✓
- Mechanical Engineer ✓
- IT Specialist ✓
- Cybersecurity Specialist ✓

### For More Information:

<https://afscsoftware.dso.mil/careers> 

Send Your Resume:

 [309SMXG.Recruiting@us.af.mil](mailto:309SMXG.Recruiting@us.af.mil)

CROSSTALK  
SPONSOR



S W E G  
S O C I A L S

