# AI4SE & SE4AI

RESEARCH AND APPLICATION WORKSHOP
SEPTEMBER 17-18, 2024

SYSTEMS ENGINEERING RESEARCH CENTER

U.S. ARMY

DEVCOM ARMAMENTS CENTER

SUMMARY

# EXECUTIVE SUMMARY

## OBJECTIVE

The US Army DEVCOM Armaments Center (AC) Systems Engineering Directorate (SED) and the Systems Engineering Research Center (SERC), a University Affiliated Research Center (UARC) for the Department of Defense (DoD), jointly sponsored the fifth Artificial Intelligence for Systems Engineering & Systems Engineering for Artificial Intelligence (AI4SE & SE4AI) Research and Application Workshop on September 17-18, 2024. The two-day event—held in person at George Mason University, Arlington, VA—gathered almost 200 participants, with nearly equal representation from government, academia, and industry, to learn from leaders using AI in this space, share ideas, and further explore outcomes that resulted from the previous AI4SE & SE4AI workshops.

### *The Workshop*

The conference theme, "Safer AI-Enabled Complex Systems: Responsible Deployment of AI through Systems Engineering," aimed to foster discussions and insights on how systems engineering (SE) can support the development of robust, efficient, and ethical AI systems, and how AI tools can support SE processes by enabling enhanced decision-making, optimization, validation, and verification. Various discussion sessions were conducted over the two-day event and this year's total of 40 presentations was almost double last year's total of 24 and these were selected from 74 submitted abstracts, an increase from 69 submitted last year.

This year's workshop elected to focus more on SE4AI, although both concepts were covered. SE4AI focuses on leveraging SE principles to develop AI enabled systems (AIES) that are safe, robust, and efficient, while extending them in response to the nature of AIES. Discussions highlighted that SE practices and processes must evolve to support AI technologies and systems that are growing rapidly in complexity. Such transformation will allow SE to guide advancements in responsible, reliable, and ethical AI, particularly as autonomy plays an increasingly larger role in existing and future systems. In the realm of AI4SE, discussions highlighted how integrating AI into SE can support the discipline's evolution and improve speed, reduce costs, and enhance the quality of designing, developing, and certifying mission-critical systems.

Day 1 of the event closed with a talk that reinforced the importance of designing and deploying AI systems that are reliable and trustworthy. A key issue was the potential mismatch between on-the-ground conditions where AI systems are deployed and the technology's performance, suggesting there is still a need for human operators to assist in addressing AI-related challenges. Using model-based systems engineering (MBSE) for validation is being explored to build trust in AI. This includes developing MBSE artifacts and infrastructure to ensure the trustworthiness of autonomous systems, which reinforced the importance of infrastructure to support existing technologies and the future advancements these will enable.

Day 2 closed with a presentation on the INCOSE AI Working Group that explores AI's relevance to SE, develops educational materials to introduce AI to the SE community, and creates INCOSE products for advancing SE4AI and AI4SE. Group meetings are held approximately every two months, featuring short talks followed by discussions, and the group is working on several full-day tutorials, an AI Systems Primer, and a joint project with the Requirements Working Group. The INCOSE Systems Engineering Journal Special Issue was introduced, along with plans for an AI Lightning Talk at IW 2024, scheduled for February 2025, in Seville, Spain. Lastly, the SERC INCOSE SE4AI and AI4SE Virtual Workshop was introduced, with the abstract submission deadline set for October 5, 2024.

### *Key Points*

- AIES is a multidisciplinary area that requires a variety of insights. The rapid and increasing complexity of AIES (AI-enabled systems) involves multiple subsystems interacting across different tasks, stakeholders, and sectors with varying levels of adaptation. The workshop exemplified that the path toward innovation and solutions begins from gathering representatives from different sectors for idea and expertise exchange.
- Ensuring AI reliability in critical situations is crucial. The growing complexity of AI systems, especially in safety- and mission-critical domains, highlights the need for good SE practices, as well as the limitations of traditional SE methods when applied to AI, where systems may change behavior based on operational experience. The risk in over-reliance on AI and the need to balance the technology's use with human expertise as autonomous systems become more complex was emphasized, as was the opportunity for SE to ensure such systems accurately interpret and act upon human commands.
- Systems engineers and the discipline play an important role in addressing AI trustworthiness. Systems engineers and the discipline have the opportunity to guide advancements in responsible AI development and ensure that AIES are trustworthy. The challenge is ensuring trusted AIES through improved SE design methodologies and tools. Future research priorities include building ecosystems of trust, measuring trust in AI, improving AI explainability, and systematic modeling of AI system components to enhance transparency, security, and compliance.
- SE processes will evolve. AI4SE introduces unique opportunities for new SE methods and tools. AI can assist the transformation of the SE discipline by making possible efficiency gains from using the technology for repetitive tasks and for accelerating other tasks. New SE4AI processes can address the need for a systematic way to predict potential undesirable behaviors of AIES and identify root causes, enabling stakeholders to devise preventive strategies and build trust.

- AIES introduce a challenging level of complexity, particularly for testing and evaluation (T&E). SE4AI presents an opportunity to rethink testing objectives—from validating fixed requirements to predicting performance in evolving environments. The goal is to refine testing as AI models evolve and view T&E as a continuous process. Human interaction with AI systems can complicate testing, but it can also provide a path to mitigate certain AIES risks. SE methods need to be designed for ongoing performance evaluation and monitoring and to simulate future threat environments, enabling early testing of system configurations, reducing risk, and increasing visibility.
- Human and AI teaming presents both challenges and opportunities. A transparent view of how people and technology interact in a system is key to building trust and driving refinement. Defining roles and responsibilities in AIES is particularly challenging as actions carried out by these systems may create uncertainty regarding accountability for the outcomes. While it is commonly prescribed to have a human in the loop, the challenge lies in determining the optimal placement of human involvement and the impact that has on performance risk tradeoffs. Increased human-AI teaming presents new challenges for workforce development.

****

The AI4SE & SE4AI Research and Application Workshop has grown in scope in its five years and each year has allowed an exchange on progress, challenges, and goals. This year's event continued to acknowledge the importance of confidence and trust in new technologies and systems, as well as the broad implications outside of the industry, academic, and government sectors. Safety remains a primary concern, with great risk passed on to the warfighter, as well as to members of communities in conflict. Acknowledgement that people make progress possible kept a continued emphasis on workforce development to ensure individuals understand their role within the larger, interconnected digital ecosystem focused on delivering reliable capabilities to the warfighter at the speed of relevance.

Based upon this year's registration information, 33% of attendees were from academia, 30% from industry, 28% from government, and 9% from FFRDCs (Federally Funded Research and Development Centers). Within this mix of sectors is where the answers and solutions can be developed. The workshop organizers and participants look forward to a sixth gathering in 2025 and continued guidance on evolving efficiently and effectively into the future.

# TABLE OF CONTENTS

# INTRODUCTION

This was the fifth Artificial Intelligence for Systems Engineering & Systems Engineering for Artificial Intelligence (AI4SE & SE4AI) Research and Application Workshop, jointly sponsored by the US Army DEVCOM Armaments Center (AC) Systems Engineering Directorate (SED) and the Systems Engineering Research Center (SERC), a University Affiliated Research Center (UARC) for the Department of Defense (DoD). The two-day, in-person event was held September 17-18, 2024 at George Mason University (GMU). The conference theme, "Safer AI-Enabled Complex Systems: Responsible Deployment of AI through Systems Engineering," aimed to foster discussions and insights on how systems engineering (SE) can support the development of robust, efficient, and ethical artificial intelligence (AI) systems, and how AI tools can transform the practice of SE. The workshop was attended by multiple Other Government Agencies (OGAs) and industry and academia affiliates, with almost 200 people present.

# WORKSHOP AGENDA STRUCTURE AND AUDIENCE

In our design of the workshop, the keynote addresses and plenary panels targeted specific themes that the workshop chairs wanted to emphasize. The program complemented technical talks with three keynotes representing the perspectives of industry, government, and academia. In addition, two plenary panels focused on core topics in SE4AI (testbeds) and AI4SE (AI supporting mundane tasks), as well as program updates from each of the sponsoring organizations (AC and SERC).

Mr. Matthew Rose (Global Public Sector Industry Principal, Snowflake; AIRC Innovation Panel member) served as industry keynote speaker for Day 1 of the two-day event. On Day 2, Mr. Daniel R. Mahanty (Division Director for Learning, Civilian Protection Center of Excellence; AIRC Innovation Panel member) served as morning government keynote speaker and Dr. Missy Cummings (Professor and Director of Mason Autonomy and Robotics Center, GMU) delivered the afternoon academic keynote. The first plenary panel, The Need for Sociotechnical System Testbeds for AI-Enabled Systems, highlighted the SE4AI theme with a discussion of the need for effective test and evaluation methods in realistic human-AI testbeds. The second plenary panel, Opportunities and Risks for Leveraging Generative AI to Support SE Processes, highlighted the AI4SE theme with a discussion of the use of generative AI in SE tasks. The workshop agenda was structured into the following two tracks along with presentations on corresponding relevant topics:

**TRACK 1 | SE4AI**

- Safety Frameworks highlighted the application of safety best practices and continuous evaluation to ensure AI systems are safe, effective, and secure;
- AI "-ilities" highlighted needs and methods to improve the quality of AI systems with respect to common "-ilities" such as safety, robustness, and trust;
- Test and Evaluation (T&E) highlighted evaluations as useful to improve internal processes and governance of AI systems and provide assurance of their trustworthiness;
- Human Autonomy Integration and Trust highlighted the importance of preserving the autonomy and privacy of humans interacting with AI systems; and
- SE Methodologies for AI Expert Systems highlighted the effective design of systems that mimic human expertise and decision making in specific domains.

**TRACK 2 | AI4SE**

- AI for Design and Governance of Complex Systems highlighted how efficiencies gained by AI-optimized system designs can allow systems engineers to focus on more complex aspects of design and analysis;
- AI at the Enterprise Level highlighted how AI can support integrating systems capabilities to support enterprise-wide modernization priorities;
- AI4SE in Digital Engineering (DE) highlighted opportunities for integrating AI and DE to guide digital transformations to maintain engineering rigor and achieve efficiencies;
- Large Language Models (LLMs) for SE Artifacts highlighted opportunities to gain benefits and efficiencies with LLM-generated artifacts; and
- AI-Aided Systems Engineering and Design highlighted exploring using AI to support creation of complex systems that are more secure and resilient throughout the lifecycle.

Each track was moderated with an interactive discussion and Q&A at the end.

Presentation materials for the entire workshop are available via the event page on the SERC website.

# WORKSHOP KEYNOTES AND PERSPECTIVES

**DAY 1 | MORNING KEYNOTE**

**Mr. Matthew Rose**, *Global Public Sector Industry Principal, Snowflake; AIRC Innovation Panel member*

Mr. Rose highlighted the potential of AI in both the military and public sectors, the challenges of market and regulatory differences, and the need for collaboration across multiple sectors. The discussion noted that small businesses can play a vital role, particularly in government contracts, and the need for new policies to handle AI's stochastic nature.

Mr. Rose drew from his experience in the military, highlighting that the Army is approaching AI as a multidisciplinary area, combining different sectors. He mentioned that many AI leaders today are data-driven companies, and that AI and machine learning (ML) have been central to advances in defense, manufacturing, and the aerospace sectors. AI's role in fraud detection, voice recognition for emergency systems, healthcare, and other applications were highlighted as vital developments.

Mr. Rose stressed that academic research could further define AI's potential use cases. He highlighted the importance of recognizing market transitions, particularly how data has shifted from being a liability to an asset. A parallel was drawn with airline programs, NASA, and public-sector applications. Mr. Rose emphasized the need for structure and governance in public sector investment, particularly in defense, to ensure alignment across global markets.

**DAY 1 | US Army DEVCOM Armaments Center Perspective**

**Mr. Edward W. Bauer**, *Director of the Systems Engineering Directorate (SED), US Army DEVCOM Armament Center (AC)*

Mr. Bauer focused on DEVCOM AC efforts to integrate AI into systems engineering (SE). DEVCOM AC's AI strategy is focused on creating trusted and assured AI products that can meet operational requirements safely and involves AI-enabled armament systems, tools and ecosystems for AI use, workforce recruitment and skill development, and partnerships with government, industry, and academia. He underlined the importance of integrating AI and SE for future Army applications, building a digital foundation, and ensuring AI's trust and reliability in critical military systems. He emphasized the need for improved data-driven decision-making, ML, and reinforcement learning to optimize processes.

DEVCOM AC is actively focusing on the opportunities that exist for newcomers in AI4SE, particularly drawing on its history of developing robust training programs to meet unique needs, continuing to leverage these, and forming partnerships. Mr. Bauer noted human and AI teaming as a significant challenge that requires ongoing research to explore and identify practical approaches.

**DAY 2 | MORNING KEYNOTE**

Mr. Daniel R. Mahanty, *Division Director for Learning, Civilian Protection Center of Excellence; AIRC Innovation Panel member*

Mr. Mahanty discussed the Center's mission to bridge the divide between policy and the work of engineers in order to understand the consequences of the technologies being developed. The presentation spoke to the current emphasis on responsible AI and recognition of the broader impacts, including the ethical implications, of this technology.

Mr. Mahanty noted two effects of AI technologies to be considered when developing policy: 1. the need to consider the effects of new technologies beyond immediate geographic zones of combat, and 2. the need to expand the definition of civilian harm to include interruptions to critical structures and systems. The goal is to mitigate these effects in all aspects of military planning. The current moment is significant for these discussions with the increasing potential for large-scale conflict and the DoD goal of enhancing capabilities, including decision making, specifically through the use of AI. The Center calls for alignment among stakeholders across sectors to develop guidelines for responsible use of new technologies and include language within the military that references such use. The US can model leadership in protecting civilians including by integrating harm mitigation action plans in all aspects of defense and looking at root causes that create technology biases that lead to harm.

Mr. Mahanty noted there is promise in the risk management potential of AI, but the risks inherent in the technology need to be considered, particularly where human-machine interaction occurs. Overconfidence in machines can affect decision making and the risk exists of adversaries altering and manipulating data. There also exist opportunities to leverage technology to protect civilians, e.g., in the field of targeting, AI technology has the potential to better identify civilians and the dynamic patterns of life in an environment. The Center is exploring whether large amounts of data lead to better models to understand dependencies of civilian life in a combat environment.

The Center is reaching out to a broader variety of stakeholders on how each can apply the guidance on risk mitigation and contribute to the larger discussion. The Center is also trying to identify performance attributes they want to see realized and is talking with defense partners on how to achieve these goals.

**DAY 2 | AFTERNOON KEYNOTE**

**Dr. Missy Cummings**, *Professor and Director of Mason Autonomy and Robotics Center, GMU*

Dr. Cummings called for SE practices to be formalized in AI systems testing and certification, along with developing a responsible AI governance framework and workforce. AI reacts well to standard behavior but its lack of intuitive knowledge results in the technology's difficulty responding to non-standard behavior. Many AI-first organizations have not done proper SE to test their systems, often relying on modeling and simulation alone to cut costs and reduce time to market. To illustrate, Dr. Cummings shared insights on key issues with autonomous vehicles (Avs) that she gained through her work as a fighter pilot with the US Navy, a researcher at MIT and Duke, and advisor to the National Highway Safety Transportation Administration (NHSTA). One issue is the misrepresentation of AV safety. As an example, Dr. Cummings pointed out that current AVs often perform better than distracted rideshare drivers but are far less safe than the average human driver. Another key issue is AV hallucination that can cause rapid deceleration. Due to these issues, AVs get rear-ended at twice the rate of traditional cars. Other issues include poor planning decisions and the inability to deal with unexpected actions.

**DAY 2 | SERC Perspective**

**Dr. Zoe Szajnfarber**, *Professor/Chief Scientist, GWU/SERC*

Dr. Szajnfarber emphasized the important role of systems engineers in ensuring that AI systems function reliably and ethically, contributing to the broader challenge of "trustworthiness" in AI. She noted that SERC is one of four founding members of the Archimedes initiative, an international partnership that aims to accelerate innovation in SE research. In its most recent workshop held in summer 2024, Archimedes participants collaborated on defining and measuring "trust" and "trustworthiness" in AI, identifying the complexities involved in ensuring safe AI, and working on responsible generative AI. The resulting discussions centered on future research priorities, such as building ecosystems of trust, measuring trust in AI, and improving AI explainability. These findings will inform future collaborations to develop a research roadmap for "trustworthy AI," highlighting the need for SE to guide advancements in responsible AI development.

# WORKSHOP PANELS

## DAY 1

**Plenary Panel | The Need for Sociotechnical System Testbeds for AI-Enabled Systems**
*Moderator:* Dr. Laura Freeman*, Virginia Tech*
Panelists: Mr. David Jin, *CDAO*; Dr. Jose Rodriguez, *DEVCOM AC Tactical Behavior Research Lab*; Dr. Zoe Szajnfarber, *GWU/SERC*; Mr. Miles Thompson, *MITRE AI Discovery Lab*

The panel gathered insights from representatives of government, industry, and academia on the challenges and opportunities of testing for AIES applied to real-world problems. While approaching the topic from diverse perspectives, panelists were in general agreement that both model and systems testing is essential for fostering system trust and managing risk, but AIES introduce a level of complexity that makes effectively designing appropriate testbeds difficult. Strategies proposed for testbeds and testing included thinking carefully about scale, scope, and interfaces to design test models that do well at approximating specific system behaviors, engaging human testers from the community of use and who demonstrate the actual behaviors of the users who interact with the systems, and establishing AI systems testing best practices and communities of practice to provide a standardized but adaptable foundation.

## DAY 2

**Plenary Panel | Opportunities and Risks for Leveraging Generative AI to Support SE Processes**
*Moderator:* Dr. Peter Beling*, Virginia Tech*
Panelists: Ms. Allison Banzon, *AI Ethics and Learning Analytics, SAIC*; Mr. Chris Schwalm, *CVP Corp*; Dr. Daniel Selva, *Texas A&M University*

The panel discussed the opportunities and risks of leveraging generative AI (GenAI) in SE processes. Highlighted were the efficiency gains from using AI for tasks such as generating documentation, developing test cases, and completing compliance checks, as well as accelerating requirement development with business analysts. However, there is risk in over-reliance on AI and it was emphasized that its use needs to be balanced with human expertise. A panelist focused on AI ethics stressed the importance of data hygiene, privacy, and maintenance of basic process elements, cautioning against sacrificing safety for speed in AI implementation.

Insights into the technical landscape were provided, noting that GenAI extends beyond language models and incorporates large foundation models with capabilities like zero-shot learning. It was suggested to fine-tune existing models like GPT-4 for SE use cases. The panelists discussed the need for benchmarking, prompt engineering, and careful evaluation before full-scale AI adoption, with a focus on transparency, cost management, and data protection. The balance between open-source and commercial AI tools was also addressed, stating that the value of both is dependent on specific needs.

## WORKSHOP PRESENTATIONS

Please refer to the event page on the SERC website to download presentation slides.

## PRESENTING ORGANIZATIONS

- Army Command, Control, Communications, Computers, Cyber, Intelligence, Surveillance and Reconnaissance (C5ISR) Center
- Army DEVCOM Aviation & Missile Center RAM Division
- Army Logistics Modernization Program
- Booz Allen Hamilton
- Carnegie Mellon University Software Engineering Institute
- CentraleSupélec, Université Paris-Saclay
- Colorado State University
- George Mason University
- George Mason University Command, Control, Communications, Computing, Cyber, and Intelligence Systems (C5I) Center
- Georgia Tech Research Institute
- Leidos
- Lockheed Martin Advanced Technology Laboratories
- ManTech
- MITRE Corporation
- NASA Langley Research Center
- National Institute of Standards and Technology
- Naval Information Warfare Center
- Northeastern University
- Purdue University
- Shoal Group
- Stevens Institute of Technology
- Strategic Ai Services
- Systems Engineering Research Center
- Tangram Flex
- The George Washington University
- The University of Alabama in Huntsville
- United States Military Academy
- University of Adelaide
- University of Maryland
- University of Southern California Information Sciences Institute
- Virginia Tech National Security Institute

# ACKNOWLEDGEMENTS

The organizers would like to express thanks to the presenters in this workshop who generously shared their knowledge, expertise, and experience. Thank you to DEVCOM AC Systems Engineering Directorate and SERC for planning and facilitating, and to all the attendees for the open discussion, ideas, and information exchange. It was again an opportunity to gather the community together to advance SE and AI.

## WORKSHOP ORGANIZERS

**Executive Host:**
Dr. Dinesh Verma, *SERC Executive Director, Stevens Institute of Technology*
Mr. Edward W. Bauer, *Director of the Systems Engineering Directorate (SED), US Army DEVCOM Armaments Center (AC)*

**Technical Committee Leads:**
Dr. Zoe Szajnfarber, *GWU/SERC*
Dr. Peter Beling, *Virginia Tech*

**Moderators:**
Dr. Peter Beling, *Virginia Tech*
Dr. Myron Hohil, *US Army DEVCOM AC*
Dr. Ali Raz, *George Mason University*
Mr. Benjamin Schumeg, *US Army DEVCOM AC*
Dr. Val Sitterle, *Georgia Tech Research Institute*
Mr. Al Stanbury, *US Army DEVCOM AC*
Dr. Ralph Tillinghast, *US Army DEVCOM AC*
Dr. Zoe Szajnfarber, *GWU/SERC*

## ACRONYM LIST

AC – Armaments Center
AIBOM – AI Bill of Materials
AIES – AI enabled systems
AI/ML – artificial intelligence/machine learning
AS – autonomy stack
AV – autonomous vehicle
C2 – command and control
DE – digital engineering
DEVCOM – Combat Capabilities Development Command
DevOps – development and operations
DNN – deep neural network
DoD – Department of Defense
DSL – domain-specific language
GenAI – generative AI
HAI – human artificial intelligence
HAT – human-AI teaming
HTM – human-technology management
LLM – large language model
LoA – level of autonomy
M2ALPS – Multi-Mode Logistics Planning System
MARL – multi-agent reinforcement learning
MBSE – model-based systems engineering
ML – machine learning
MOE – measure of effectiveness
OGA – other government agencies
QA – quality assurance
R3+ – robustness, resilience, reliability, and antifragility
RAG – retrieval augmented generation
SAG – synergistic adaptive governance
SE – systems engineering
SED – Systems Engineering Directorate
SEPTAR – Systems Engineering Processes to Test AI Right
SERC – Systems Engineering Research Center
SMoCLib – semantic model component library
SoS – system of systems
STPA – systems theoretic process analysis
SysML – systems modeling language
T&E – testing and evaluation
UAF – unified architecture framework
UARC – University Affiliated Research Center
UAV – unmanned autonomous vehicle
V&V – verification and validation