AI4SE & SE4AI

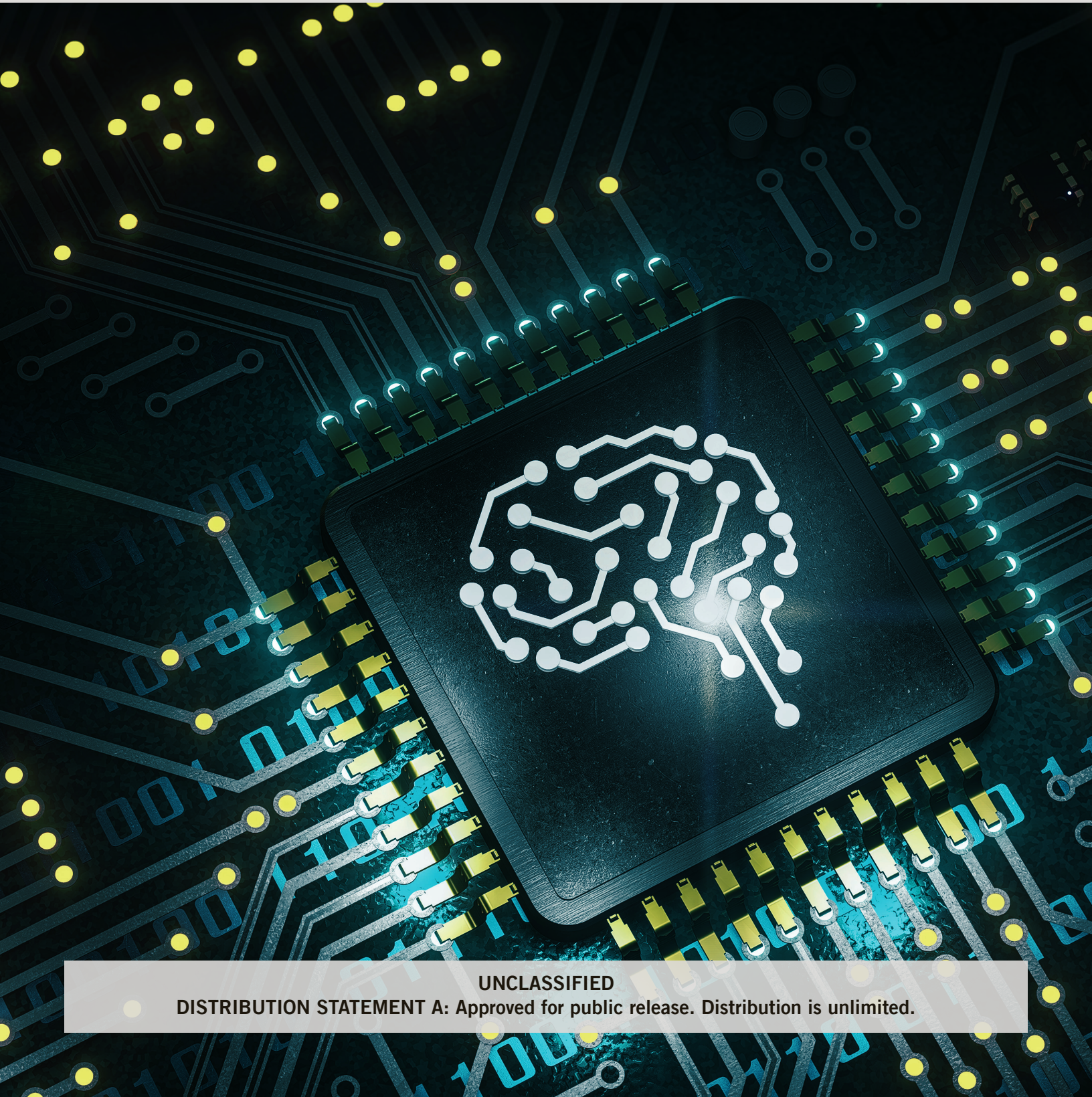
RESEARCH AND APPLICATION WORKSHOP
SEPTEMBER 21-22, 2022

Sponsored and organized by DEVCOM Armaments Center (AC) Systems Engineering Directorate (SED) and the Systems Engineering Research Center (SERC)

SYSTEMS ENGINEERING RESEARCH CENTER

U.S.ARMY

DEVCOM ARMAMENTS CENTER



UNCLASSIFIED

DISTRIBUTION STATEMENT A: Approved for public release. Distribution is unlimited.

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 third Artificial Intelligence for Systems Engineering & Systems Engineering for Artificial Intelligence (AI4SE & SE4AI) workshop on September 20-21, 2022. The two-day hybrid event—held in person at Stevens Institute of Technology, Hoboken, NJ, and online via ZoomGov—gathered participants from government, academia and industry to learn from leaders using AI in this space, share ideas and further explore outcomes that resulted from the two previous AI4SE & SE4AI workshops.

The Workshops

A total of 17 presentations, two panels and various discussion sessions were conducted over the two-day event and focused on topics within the areas of Artificial Intelligence for Systems Engineering (AI4SE), Systems Engineering for Artificial Intelligence (SE4AI), AI4SE and Development (DEV), and SE4AI and Verification and Validation (V&V). A unifying theme was *the continued need for organizations and systems to be agile and keep pace with the dynamic nature of AI toward the goal of delivering the most relevant and effective tools to the soldier in the field*. Areas identified as essential in prior meetings maintained a high level of significance throughout this year's workshop, notably: the importance of the reliability of new systems to build user confidence and maintain global competitive advantage; and the need to update business processes, particularly workforce development and retention. Discussions consistently highlighted the importance of data, particularly its acquisition, analysis, maintenance, reliability, and security; its role in human and machine learning; and its role in the testing and validation of AI-enabled systems to mitigate risks and build trust.

Outcomes

The previous years' workshops guided the selection of presentations, panels and discussions of this 2022 event. As this year's workshop sought to address roadblocks to continued progress,

the following themes evolved:

- The topic of AI4SE & SE4AI and associated endeavors are complex: Each of the 2022 workshop's four tracks explored the complexity of human-machine teaming and of the associated endeavor being undertaken by systems engineering. Presentations and discussions considered how AI can augment and enhance SE processes, what SE can do to address the limitations of AI, and what are the particular challenges that AI presents to the SE discipline. V&V and testing and evaluation (T&E) of new systems need to continue evolving to better understand the tests being used and the future use of the systems.
- AI is a system: The task of mitigating the uncertainty of AI and providing guardrails against emergent events needs to be shared by designers, requirement setters and systems engineers and needs to begin from the start of system design. Testing of AI needs to be done in the context of a system; AI provides the benefit of automating certain processes, while systems can be shaped to recognize and compensate for the limitations of the AI. Powerful tools are currently available, and more are being developed to address the particular needs of the SE community.
- There are opportunities and challenges as AI4SE and SE4AI come together: Digital twins continue to be indispensable tools for managing the lifecycle of AI systems and to the systems engineers when designing those systems. The digital representation of a physical asset allows for real-time learning and testing that enhance system refinement and maintenance and build critical trust.
- Expanded thinking around training is critical: Investment in training, both for the users and the developers of systems with AI, must continue to be a focus. New AI capabilities, enabled and enhanced by validated and actionable data, can train warfighters in use of the next generation of systems as well as those who create the tools to deal with systems that rely on AI. Training must go beyond data scientists and the software engineers and must include the interdependencies of emergent systems. Hands-on, interdisciplinary training that helps individuals understand their role within the digital

ecosystem can help the US catch up to other global players. The engineers and the warfighter must arrive in the future together.

- Navigate cultural and policy roadblocks: AI-based applications will continue to permeate future SE and acquisition practices. AI applications are fundamentally digital (data and software) and will evolve with the digital transformation of government policy and practice. Services will have to share data, integrate with each other, and reconsider their siloed cultures and business processes. The Acquisition Innovation Research Center (AIRC) continues to research policies related to new technologies and whether they drive or inhibit progress.

The AI4SE & SE4AI Research Workshop has grown in attendance and scope in its three years; the panels and presentations clarify how AI and SE are co-evolving and will continue to do so. This year, as in the two preceding years, the workshop allowed an exchange on the digitization of SE, the accompanying technical and cultural shifts, and the integration of diverse fields and people. More powerful machines and an increase in data will continue to drive the proliferation of AI. Unprecedented technology generates uncertainty. Any loss of confidence and trust in new technologies and systems will set back progress. Safety is a primary concern, with the greatest risk passed on to the warfighter.

This year's workshop highlighted the role of data in developing the AI- and ML-enabled tools that allow everyone, from the systems engineer to the warfighter, to do things they could not accomplish previously. While AI and automation can support productivity and efficiency, it is people who develop and adopt these tools. People need the support to understand their role within the larger, interconnected digital ecosystem that is focused on the shared goal of delivering capabilities to the warfighter at the speed of relevance. Increasingly, systems engineers need to think of the problems to be solved and the opportunities to be harnessed.

The mix of industry, academia and government represented at the annual workshops is where the answers and solutions lie. The workshop organizers and participants look forward to a fourth conference in 2023 and continued guidance on evolving into the future efficiently and effectively.

Table of Contents

INTRODUCTION

This workshop was the third Artificial Intelligence for Systems Engineering & Systems Engineering for Artificial Intelligence (AI4SE/SE4AI) Research 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 objective of the workshop was to discuss and define Systems Engineering (SE) and Artificial Intelligence (AI) challenges, areas of exploration and methodologies to use, and ways in which to collaborate and research in the upcoming years***.

WORKSHOP AGENDA STRUCTURE AND AUDIENCE

DEVCOM AC and the SERC jointly hosted the two-day hybrid event, held in-person at Stevens Institute of Technology, Hoboken, NJ, and online at ZoomGov. On Day 1, Mr. R. Chris DeLuca (Director, Specialty Engineering OUSD(R&E)/ED(SE&A)) served as Keynote Speaker and provided relevant perspectives for both days' sessions. The workshop was attended by multiple Other Government Agencies (OGA) and Industry and Academia affiliates with over 200 people registering and attending the hybrid workshop. The workshop agenda was structured into the following four (4) tracks:

- **Artificial Intelligence for Systems Engineering (AI4SE)**
- **Systems Engineering for Artificial Intelligence (SE4AI)**
- **AI4SE and Development (DEV)**
- **SE4AI and Verification and Validation (V&V)**

Each track had 4-5 presentations on relevant topics, and audience members were able to collaborate and ask questions throughout the briefings in person and via the virtual chat feature. Each session was moderated with an interactive discussion and Q&A at the end.

Presentation materials for the entire workshop are available via the SERC website: https://sercuarc.org/event/ai4se-and-se4ai-workshop-2022/

WELCOME & OPENING REMARKS

DAY 1 - Dr. Anthony Barrese, *Dean, School of Systems & Enterprises, Stevens Institute of Technology*

Dr. Jason Cook, *SSTM for SE Research*, *US Army DEVCOM Armaments Center*

Attendees were welcomed and it was noted the 2022 event was the first time the jointly sponsored workshop was presented in a hybrid format (in-person at Stevens Institute of Technology and online at ZoomGov). Since its inauguration in 2020, the workshop continues to provide an opportunity for exchange among and contributions by representatives of Government, Academia and Industry toward continued refinement and rich application of artificial intelligence (AI) and machine learning (ML). The workshop is also an annual opportunity for stakeholders to "refresh" in a field that is moving quickly and address how to ensure safety, reliability and trust in human-machine interactions.

WORKSHOP KEYNOTE

DAY 1 | DoD Is a Data Centric Organization

Mr. R. Chris DeLuca, *Director, Specialty Engineering OUSD(R&E)/ED(SE&A)*

Data, enhanced by AI, is key to advancing the National Defense Strategy. Data sets for AI training and algorithmic models will increasingly become the DoD's most valuable digital assets. As the Department modernizes and integrates AI technologies into joint warfighting, generating DoD-wide visibility of and access to these digital assets will be vital in an era of algorithmic warfare. Sustainment attributes need to be integrated early into system design and a supportive infrastructure is needed to engineer data that will advance the Department's goals. The increasing complexity and challenges of a data-centric vision require a workforce skilled in understanding how to acquire and use the data in order to become part of an overall warfighting platform; as a result, a broad job space is opened for those in systems engineering (SE) who care about AI and how to apply it to their practice. Mr. DeLuca closed emphasizing that AI enables delivery of safe, reliable data to the warfighter, and the technology and systems being developed today are critical to the ability to protect the nation, sea lanes and world from threat.

WORKSHOP PANELS

DAY 1 PANEL | SE Workforce Development Needs

Moderator: Dr. Jason Cook, *SSTM for SE Research*, *US Army DEVCOM Armaments Center*

Panelists: Dr. Rosa Heckle, *MITRE Corporation;* Dr. Zoe Szajnfarber, *George Washington University;* Jaime Hernández-Cordero, *DoD*; Rhonda Maus, *DAU*

The evolving future of combat requires systems that are integrated, augmented, trusted, collaborative, and have the capacity to learn. Achieving modernization ahead of adversaries requires the combined efforts and contributions of the engineers who design the complex systems and the warfighters who use the resulting technologies. The challenges are numerous—from how to verify and validate systems with emergent properties to how to capture the data needed to inform system learning with cyber security still enabled—and require new skills from the workforce.

Panelists discussed key learning areas on which to focus, including understanding: the fundamentals of AI/ML sufficiently to evaluate their effect on system performance; systems architecture, what functions may be impacted by AI, and how to control critical functions; configuration management; verification and validation, and testing and evaluation (of system performance AND learning); and how to develop relevant training data sets. Individuals within an organization need to understand how their roles support the AI ecosystem. Training, education and certification were also addressed, with emphasis on hands-on, interdisciplinary approaches.

DAY 2 PANEL | Government Visions and Needs

Moderator: Dr. Ralph C. Tillinghast, *SSTM for Future Intelligence, US Army DEVCOM Armaments Center*

Panelists: Ms. Carol Pomales, *MITRE Corporation*; Dr. Raj Iyengar, *US Nuclear Regulatory Commission*; Mr. R. Chris DeLuca, *Director, Specialty Engineering*

OUSD(R&E)/ED(SE&A)

The panel discussed exploring the AI space to discover where the Services are in terms of readiness for AI, identify the desired future state, and take action towards those goals. The Undersecretary for Research and Engineering cited "trusted AI" among the top research priorities. The challenge for the T&E community is to mitigate mission risks before and after AI-enabled systems are deployed. In partnership with the T&E community, DTE&A has delineated a perspective of where to go based on its research, showing a future state of T&E of AI through analyses of six key areas: Policy, Measures, Data, User Engagement, Cybersecurity, and Infrastructure.

Continuing with a focus on how the Department, Services and divisions need to be prepared for the future, the panel also discussed digital twins and the technology's potential to enhance AI. Data allows the digital twin to enhance and upgrade the corresponding physical asset in real time; toward this use, considerations include: what kind of sensors are needed to gather and send data from the physical asset to the digital twin; the kind of data analytics needed; and the range of applications and uses for the data. Challenges and gaps include building the necessary infrastructure to enable the technology and for effective user interfaces and data analytics, and ensuring trustworthiness. There is also a need to leverage available and emerging information for wide use, for example through repositories, workshops, reports, and newsletters.

WORKSHOP PRESENTATIONS – DAY 1

TRACK 1: AI4SE

Understanding Datasets: Seeing the Unseen through Graph Automations

MITRE Corporation

An often overlooked, critical phase of implementing an AI model is the engineering used to create and understand a normalized Unified Data Fabric (UDF) from disparate data sources.

Such a UDF allows the data scientist to understand the data and its structure from multiple perspectives, increasing the likelihood of success in modeling, and is then the precursor to an adaptable AI infrastructure. During this topic, work was presented that proposes a solution that simplifies the design and construction of a UDF through the use of autogenerated components, not just for the UDF but also for the AI infrastructure.

Semi-Automated Development of Textual Requirements: Combined Natural Language Processing and Multi-Domain Semantic Approach

University of Maryland, College Park

Modern engineering systems are nearly always designed, built, operated, and maintained by teams of people and automated procedures for decision making over extended periods of time. The ability to write and manage textual requirements early in the system development life cycle is a key element in making systems that work. The presenter presented a framework that employs natural language processing (NLP) and template matching for the semi-automated development and validation of textual requirements to improve the quality of communication among disciplines throughout the system life cycle. The presentation described the capabilities of a software prototype developed for textual requirement sentences covering project cost and schedule, building foundation, construction site, equipment, and workers.

Reinforcement Learning for Autonomous and Novel Behavior in OneSAF: Potential for New TTPs, Technology Evaluation, and Requirement Definitions

Quansight

Existing semi-autonomous behavior models in the One Semi-Automated Forces (OneSAF) simulator are rule-based algorithms incapable of novel behavior beyond human input. In contrast, a new Reinforcement Learning (RL) model makes its own decisions in real time with minimal human input. Extending OneSAF via advanced artificial intelligence (AI) will open the

door for new Tactics, Techniques, and Procedures (TTPs), evaluation of new technologies, and reshape how the DoD defines requirements of the future force. The RL algorithm connects to OneSAF via a new Python API, which is independent of the RL algorithm and allows OneSAF to be controlled via Python code. A proof-of-concept model demonstrated novel behavior in navigation, survival, and firing, highlighting the potential for generating new battlefield strategies with minimal human bias and helping warfighters learn in a more creative, stimulating and instructive environment.

Natural Language Processing and Knowledge Engineering to Extract Models from Text

Stevens Institute of Technology

The presenter presented a study in the fast-growing domain of natural language processing (NLP)/Understanding. At the 2021 workshop, the presenter presented research based on extracting formal knowledge representations from text, introducing a proof of concept for one of the possible approaches. At this year's event, a comparison between the two main approaches: the symbolic and the machine learning (ML) was presented. Preliminary results show the symbolic approach seems to provide good results with lower data and a more narrow/specific but static domain. The ML approach provides good results when there is a large availability of data, and the domain is generic. The research also supported development of a knowledge representation model (the "room theory") that organizes the remembered information into an individual/domain-specific view.

The Design of Swarm Experiments: A Systems Engineering Approach to Swarm Research

US Army DEVCOM Armaments Center

While research in swarm intelligence, swarm algorithms, and swarm performance optimization is becoming prevalent in the literature, a general, formal method for selection of swarm characteristic distributions does not appear to be available. Presented were the general steps

and examples of a method that applies the Design of Experiments (DOE) methodology to swarm performance research, with a focus on research using computer models to simulate swarm behavior. The proposed Design of Swarm Experiments (DSE) method provides a mathematical way to select optimal distributions of characteristics across a swarm. DOE techniques intelligently sample the design space and ML models are used to predict the outputs of interest as a function of swarm characteristics.

TRACK 2: SE4AI

Bayesian Methods for T&E of Learning Based Systems

Virginia Tech

Presented was practical application of a new approach based in systems engineering (SE) and statistical methods to address the shortcomings of applying traditional systems that employ black-box test and evaluation (T&E) methods to complex and dynamic learning-based systems (LBS). An overarching objective of this research is to assess the risk associated with deployment of LBS using Bayesian networks, which provide a principled, probabilistic methodology for tracking performance as scenarios and systems evolve. Combining Bayesian methods with systems theoretic mechanisms used to determine equivalence and with empirical data enables the framing of a risk profile associated with LBS deployment. The research leveraged past work on the Silverfish Testbed (a networked munition system designed to engage the enemy using ground-based weapons within a denied area) and provided a vision of near-future T&E that will provide the agility to deliver LBS to the warfighter with high confidence.

NeuralSat: A Constraint Solver for Deep Neural Network Verification

George Mason University

NeuralSAT was presented as a new satisfiability (SAT) solving approach to Deep Neural Network (DNN) verification that aims to both prove and disprove properties efficiently. Inspired by the belief that constraint-based approaches can do much better, the work aims to integrate clause learning (to improve performance by addressing SAT challenges such as backtracking) in modern SAT solving with an abstraction-based theory solver (to quickly check for infeasibility) in SMT solving and abstraction-based DNN verification. NeuralSAT appears to perform better and faster than abstract-based approaches in disproving properties, suggesting the approach is effective at finding counterexamples. Initial investigation suggests that the synergistic combination of clause learning and abstraction solver is the "secret sauce" distinguishing NeuralSAT; it is hoped that with additional optimizations, NeuralSAT will set a new standard and open new research directions in DNN verification.

Test and Evaluation of AI Systems with Explainable AI and Counterfactuals

George Mason University

Current SE design, test and evaluation (DT&E) approaches are not sufficient for complex systems and system-of-systems that integrate machine learning based AI components, as the machine learning (ML) algorithms are primarily data-driven with opaque decision-making constructs. Presented were three approaches that can be added to the SE tool set to address the challenge of test and evaluation (T&E) of systems with AI components: 1. System Identifiability aims to construct an iterative experimental regime for evolving the AI development throughout the system life cycle; 2. Explainable AI (XAI) aims to transform the "black box" models of ML components to models that are understandable and interpretable by humans; and 3. Counterfactual Testing at both the component and system level investigates the what-if scenarios to understand patterns of causality in the underlying model that would otherwise be left unexposed. An example of an engineered system with classical control and Reinforcement Learning (RL) based control was presented as a contrast to the DT&E of systems with ML-based components.

How Wrong Is Wrong? Richer Characterization of AI Classification Error for Real-World Applications

George Washington University

Currently, evaluations of image classification algorithms are relatively divorced from realistic evaluations of system requirements and don't capture simple ideas of modeling how bad (for system performance) are different types of errors. The research team developed an approach that is explicit about the costs of different types of errors in the context of image classification systems where the labels of interest have a known organization or hierarchy of the class labels, which provides a potential definition for "how wrong are the wrong answers". The case study presented shows how ideas for SE can be integrated into the training and evaluation of a classifier in ways that make that classifier more robust in real-world conditions, which is an

important precursor to building trust. In future work, the idea of "how wrong an answer is" could be integrated as part of the loss function when training the network, which would more accurately connect to the potential use case of the image classifier, and has the potential to encourage the network to learn more semantically relevant features.

A framework for Certification of Continuous Learning Systems (CAI – Certification of Continuously Learning Artificial Intelligence)

NAVAIR

Presented was the Certification of Continuously Learning Artificial Intelligence (CAI) project that aims to accomplish predictive monitoring and control of continuously learning systems through implementing a hierarchical AI structure with the end goal of certifying these systems for deployment. Learning systems, particularly continuous learning systems, are able to solve problems that are difficult or impossible to delegate to traditional AI systems. The use of learning systems in military applications requires strong capabilities to accurately monitor, control and certify learning systems. Improper monitoring can lead to issues such as catastrophic forgetting (the tendency of a learning system to forget previously learned information when learning new information). To address such issues, the research team designed a trustworthy learning architecture and is in the process of performing experiments to prove its inherent safety and learning properties. CAI is currently being tested with simplistic, synthetic data with the aim of testing the architecture with real-world data sets collected from Navy carriers.

WORKSHOP PRESENTATIONS – DAY 2

TRACK 1: AI4SE & DEV

Specifying Data Requirements for AI Systems

Collins Aerospace

As artificial intelligence- (AI) and machine learning- (ML) based subsystems and components gain prevalence in complex systems, the importance of training data is driving a new specialty within the SE discipline: data requirements engineering—necessary to prevent serious AI failures often traced to relatively simple problems in the engineering of the data used to train the system. The presentation provided a closer look at the role of data in AI systems and how systems engineers will need to learn enough about the environment in which the system will operate and about the situations it could encounter to fully specify the data required to successfully train it. The framework of effective requirements engineering was used to show how data requirements can be elicited, analyzed, validated, and implemented in the data design.

Digital Assistance for Systems Requirement Discovery and Analysis Using Machine Learning Natural Language Processing Algorithm

NASA Langely Research Center

The complexity of Urban Air Mobility (UAM) airspace evolution requires a planning tool to effectively organize, integrate and communicate the research that will guide the evolution of UAM operations in the National Airspace System (NAS). Presented as such a tool was the UAM airspace research roadmap (Roadmap), being developed as a new SE methodology that leverages model based systems engineering (MBSE) and machine learning natural language processing (ML NLP, or just NLP) capabilities. This novel practice revolutionizes requirement discovery and analysis by which an expert user team, consisting of UAM SE and the NLP

software expert, uses these documents to train the Doc2Vec model (an NLP algorithm based on a single-hidden layer neural network) through tuning several hyperparameters. After validation of the newly trained model, the team can package this add-in application and deploy it to the UAM SEs as well as general users, both of whom can then install and use the application to search through the UAM body of knowledge efficiently.

Tactical Ammunition Management Micro Services (TAMMS) – Human-Machine Teaming by Cognitive Intelligent Agents Providing Sustainment Decision Support

US Army DEVCOM Armaments Center

Presented was an ongoing effort to utilize intelligent agents as a virtual staff to manage the sustainment and supply chain of supported military formations as their weapon systems consume ammunition. These intelligent agents are a specific form of AI that enables adaptive planning and execution management. To collaborate, these agents all work on knowledge graphs specifically curated with semantically interconnected data that is accessible to both the agents working on the plan and the human decision makers. Additionally, these plans and other details from the knowledge graphs are made accessible through microservices and other interfaces. The overall goal is to utilize these sustainment reasoners on incoming platform ammunition inventory data from field artillery formations in order to have them generate ammunition consumption trends, an ammunition storage supply posture and transportation plans for sustainment throughout the supply chain.

Computer Vision Based Machine Learning for Deck Operations

NAVAIR

Presented were two ongoing research projects, PETA (Pose Estimation and Tracking for ASIST) and PATRIOT (Panoramic Asset Tracking of Real-time Information for the Ouija Tabletop), that leverage computer vision-based ML components to improve the situational awareness and

effectiveness of users. PETA creates a landing assistant tool to estimate depth of helicopters for recoveries without hardware modifications to the ships on which it will be deployed. PATRIOT creates a Ouija-board interface that can estimate the position and orientation of aircraft, people and support equipment on the flight deck. The goal is to create a proof of concept in simulation using the same core algorithms for both projects. Multiple realistic simulations have been created to generate data to train and test the algorithms, which have accurately extracted the pose of aircraft and people in simulation meeting the main requirements. Future steps include the collection of additional real-world data, annotation of the data, and retraining the algorithms to ensure a successful transition to the fleet. Anticipated benefits to the warfighter include reducing cognitive workload and manning requirements, collecting data to improve logistics, and enabling an automation gateway for future efforts to improve efficiency and safety.

TRACK 2: SE4AI & V&V

Towards the Use of Deep Learning Neural Networks for System Validation Testing of Tightly Coupled Complex Systems

George Mason University

Systems engineers face the challenge to develop Validation Test Plans and conduct Validation Testing to ferret-out the scenarios in which the interactions, over time, between tightly coupled system components result in migration of a component into a hazardous operating regime. This task must be completed before such interaction occurs in the field. Presented was an approach using Deep Learning Neural Networks (DLNN) to generate System Validation Test Scenarios that try to reach a complete set of combinatorics of initial conditions and time-dependent trajectories for such systems. The underlying principle for the success of this approach lies in the ability of the DLNN to "learn" the behavior of the component interactions over time, and apply that "learned model" to combinatorics that are not part of the Training/Testing data set. The characteristics of the systems for which the DLNN can learn the

underlying behavior is a topic of ongoing research.

Uniting Hierarchical Planning and Model-Based Systems Engineering to Automate Failure Recovery Planning

Adventium Labs

Crewed and uncrewed aerospace systems are increasing in complexity and decreasing the reliance on ground-based operators. Failure Recovery Instruction Generation using Automata derived from Traditional Engineering (FRIGATE) models combine planning and system models and automatically perform the necessary failure recovery plan verification and generation for such systems more rapidly, accurately and cost effectively than could be performed manually. Presented were the semantics and use of the Goal Function Tree (GFT) that the FRIGATE team updated to accommodate model checking analysis.

Enabling SE for AI with Test and Evaluation Harnesses for Learning Systems

Virginia Tech

A lack of best practices for testing and evaluation (T&E) of learning systems hinders meeting the demand for operational uses of ML. Presented was a framework for best practices, described as T&E harnesses, that corresponds principally to the task of engineering a learning system, placing T&E for ML into the broader scope of SE processes. The presentation drew from recent findings in experimental design for ML, combinatorial interaction testing of ML solutions and the general systems modeling of ML. Discussed was the conclusion that existing best practices for T&E form a subset of what is needed to rigorously test for system-level satisfaction of stakeholder needs.

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 yet again an opportunity to bring the community together.

APPENDIX

WORKSHOP ORGANIZERS

Executive Hosts:

Dr. Dinesh Verma, *SERC Executive Director, Stevens Institute of Technology*

Dr. Jason Cook, *SSTM for SE Research, US Army DEVCOM Armaments Center*

Workshop Leads:

Mr. Tom McDermott, *SERC – Stevens Institute of Technology*

Ms. Kara Pepe, *SERC – Stevens Institute of Technology*

Mr. Albert Stanbury, *US Army DEVCOM Armaments Center Systems Engineering Directorate*

Moderators:

Dr. Jason Cook, *SSTM for SE Research, US Army DEVCOM Armaments Center*

Dr. Ralph C. Tillinghast, *SSTM for Future Intelligence, US Army DEVCOM Armaments Center*

ACRONYM LIST

AIRC – Acquisition Innovation Research Center

AI/ML – Artificial Intelligence/Machine Learning

CAI – Certification of Continuously Learning Artificial Intelligence

CCDC AC – Combat Capabilities Development Command Armaments Center

DoD – Department of Defense

DE – Digital Engineering

DNN – Deep Neural Network

FRIGATE – Failure Recovery Instruction Generation using Automata derived from Traditional Engineering

MBSE – Model-Based Systems Engineering

NLP – Natural Language Processing

OGA – Other Government Agencies

OneSAF – One Semi-Automated Forces

SE – Systems Engineering

SED – Systems Engineering Directorate

SERC – Systems Engineering Research Center

SHAP – Shapley Additive Explanations

SSTM – Senior Scientific Technical Manager

SysML – Systems Modeling Language

TAMMS – Tactical Ammunition Management Micro Services

T&E – Testing and Evaluation

TTP – Tactics, Techniques and Procedures

UARC – University Affiliated Research Center

V&V – Verification and Validation

XAI – Explainable Artificial Intelligence

RESEARCH AND APPLICATION WORKSHOP

AI4SE & SE4AI

2022 SEPTEMBER 21

DAY 1 SCHEDULE
(Bissinger Room)

9:00 AM - 10:00 AM
Check-in / Networking

10:00 AM - 10:15 AM
Welcome & Opening Remarks
Dr. Anthony Barrese
Dean, School of Systems & Enterprises
Stevens Institute of Technology
Dr. Jason Cook
US Army DEVCOM Armaments Center

10:15 AM - 11:00 AM
Keynote
Mr. R. Chris DeLuca
Director, Specialty Engineering
OUSD(R&E)/ED(SE&A)

11:00 AM - 12:15 PM
PANEL: Workforce Development
Dr. Jason Cook *(moderator)*
US Army DEVCOM Armaments Center
Dr. Rosa Heckle
MITRE Corporation
Dr. Zoe Szajnfarber
George Washington University
Jaime Hernández-Cordero
DoD
Rhonda Maus
DAU

12:15 PM - 1:00 PM
Lunch / Networking

1:00 PM - 4:45 PM
Track 1 and Track 2

4:50 PM - 5:15 PM
Day 1 Closing Remarks

5:15 PM - 6:30 PM
Reception (Babbio Atrium)

TRACK 1: AI4SE
(HOWE 102)

1:00 PM - 1:30 PM
The Importance of a Unified Data Fabric for AI Modeling
MITRE Corporation

1:30 PM - 2:00 PM
Semi-Automated Development of Textual Requirements: Combined Natural Language Processing and Multi-Domain Semantic Approach
University of Maryland, College Park

2:00 PM - 2:30 PM
Reinforcement Learning for Autonomous and Novel Behavior in OneSAF: Potential for New TTPs, Technology Evaluation, and Requirement Definitions
Quansight

2:30 PM - 3:00 PM
AI4SE Track Discussion

3:20 PM - 3:50 PM
Natural Language Processing and Knowledge Engineering to Extract Models From Text
Stevens Institute of Technology

3:50 PM - 4:20 PM
The Design of Swarm Experiments: A Systems Engineering Approach to Swarm Research
US Army DEVCOM Armaments Center

4:20 PM - 4:45 PM
AI4SE Track Discussion

TRACK 2: SE4AI
(HOWE 104)

1:00 PM - 1:30 PM
Bayesian Methods for T&E of Learning Based Systems
Virginia Tech

1:30 PM - 2:00 PM
NeuralSat: A Constraint Solver for Deep Neural Network Verification
George Mason University

2:00 PM - 2:30 PM
Test and Evaluation of AI Systems with Explainable AI and Counterfactuals
George Mason University

2:30 PM - 3:00 PM
SE4AI Track Discussion

3:00 PM - 3:20 PM Break

3:20 PM - 3:50 PM
How Wrong is Wrong? Richer Characterization of AI Classification Error for Real-World Applications
George Washington University

3:50 PM - 4:20 PM
CAI - Certification of Continuously Learning Artificial Intelligence
NAVAIR

4:20 PM - 4:45 PM
SE4AI Track Discussion

4:50 PM - 5:15 PM Day 1 Closing Remarks

5:15 PM - 6:30 PM Reception (Babbio Atrium)

SYSTEMS ENGINEERING RESEARCH CENTER

RESEARCH AND APPLICATION WORKSHOP

AI4SE & SE4AI

2022 SEPTEMBER 22

DAY 2 SCHEDULE
(Bissinger Room)

8:30 AM - 9:00 AM
Check-in / Networking

9:00 AM - 9:15 AM
Welcome & Morning Remarks

9:15 AM - 10:45 AM
PANEL: Government Visions and Needs

Dr. Ralph C. Tillinghast *(moderator)*
US Army DEVCOM Armaments Center

Ms. Carol Pomales
MITRE Corporation

Dr. Raj Iyengar
US Nuclear Regulatory Commission

Mr. R. Chris DeLuca
Director, Specialty Engineering
OUSD(R&E)/ED(SE&A)

10:45 AM - 11:00 AM
Networking / Break

11:00 AM - 2:40 PM
Track 1 and Track 2

2:45 PM - 3:00 PM
Workshop Closing Remarks

TRACK 1: AI4SE & DEV
(HOWE 102)

11:00 AM - 11:30 AM
Specifying Data Requirements for AI Systems
Collins Aerospace

11:30 AM - 12:00 PM
Digital Assistance for Systems Requirement Discovery and Analysis Using Machine Learning Natural Language Processing Algorithm
NASA Langley Research Center

12:00 PM - 12:30 PM
AI4SE & DEV Track Discussion

1:15 PM - 1:45 PM
Tactical Ammunition Management Micro Services (TAMMS) - Human-Machine Teaming by Cognitive Intelligent Agents Providing Sustainment Decision Support
US Army DEVCOM Armaments Center

1:45 PM - 2:15 PM
Computer Vision Based Machine Learning for Deck Operations
NAVAIR

TRACK 2: SE4AI & V&V
(HOWE 104)

11:00 AM - 11:30 AM
Towards the Use of Deep Learning Neural Networks for System Validation Testing of Tightly Coupled Complex Systems
George Mason University

11:30 AM - 12:00 PM
Uniting Hierarchical Planning and Model-Based Systems Engineering to Automate Failure Recovery Planning
Adventium Labs

12:00 PM - 12:30 PM
SE4AI & V&V Track Discussion

12:30 PM - 1:15 PM Lunch & Networking

1:15 PM - 1:45 PM
Enabling SE for AI with Test and Evaluation Harnesses for Learning Systems
Virginia Tech

2:15 PM - 2:40 PM AI4SE & SE4AI Workshop Discussion

2:45 PM - 3:00 PM Workshop Closing Remarks

SYSTEMS ENGINEERING RESEARCH CENTER