AFRL FIGHT'S ON!

THE AIR FORCE RESEARCH LABORATORY

ISSUE 79
FALL 2026

711 HPW / RHW
WARFIGHTER INTERACTIONS AND READINESS DIVISION

USSF









www.afresearchlab.com/FightsOn

U.S. AIR FORCE

TABLE OF **CONTENTS**

4 **Division Introduction**

Product Lines

6 **Readiness**

13 **Warfighter-Machine Integration**

15 **Intelligence, Analytics & Sensemaking**

16 **Modeling & Simulation Integraton Lab**

Core Technical Competencies

17 **Human Learning & Cognition**

19 **Digital Models of Cognition** CRA
Holistic Models for Decision-Making
Multi-Sensory Human System Modeling and Evaluation
Information Mastery in Cognitive Warfare

27 **Learning & Operational Training** CRA
Co-Learning for Adaptive Human-Machine Teams
Warfighter Learning Technologies
Gaming Research Integration for Learning Laboratory®

40 **Warfighter Interfaces & Teaming**

42 **Distributed Teaming & Communication** CRA
Dynamic Team Performance Assessment
Team Optimization and Recovery

50 **Human-Machine Interactions** CRA
HMI-enabled Decision Superiority
Rapid Joint Cognitive Awareness



Graphic Designer **Mr. Will Graver** (SierTek, Ltd)

Editor **Dr. Megan Morris**, 711 HPW/RHWE

U.S. Air Force photo by **MSgt. Kenneth W. Norman**

RHW

WARFIGHTER INTERACTIONS AND READINESS DIVISION

The Division is committed to advancing warfighter capability and future operational readiness by collaborating across directorates and research disciplines to enhance the cognition, resilience, and decision advantage of our Airmen and Guardians. Through focused research and applied innovation, RHW directly supports the AFRL mission to discover, develop, and deliver war-winning science and technology, transforming decision superiority into decisive air and space dominance to win today and into the future.

To view more *Fight's On!* issues, visit:

www.afresearchlab.com/FightsOn





Mr. Dean "Bear" Berry
Division Chief, 711 HPW/RHW

RHW

DIVISION INTRODUCTION

It has been another amazing year, filled with unprecedented challenges to include an AFRL restructure and funding cuts. Again, the RHW team meets these challenges with determination and conviction! Amid uncertainty the Division continued supporting the user community and transitioning capabilities. I continue to be excited and humbled to be part of the extremely talented team that comprises AFRL's Warfighter Interactions & Readiness Division, RHW.

RHW remains at the tip of the spear for addressing all aspects of Human-Machine Teaming and Integration for Decision and Information Supremacy. Another RHW focus area is Operational Test and Training Infrastructure (OTTI) leveraging our extensive Modeling, Simulation, and Analysis (MS&A) capabilities, especially our virtual systems and the Joint Simulation Environment (JSE) in a box for research.

In line with AFRL's Human Effectiveness Directorate's (RH) enduring mission of enabling, sustaining, and enhancing multi-domain capable warfighters and integrated operations, RHW directly impacts the entirety of AFRL. RHW enables a more informed, agile, and lethal force by delivering revolutionary capabilities and enhancing preparation and mission execution across the full range of military operations. RHW continues its history of collaboration across the Department of War (DoW), national laboratories, international partners, academia, industry, and AFRL to build/maintain critical relationships to address warfighters' needs and gaps. RHW leverages these synergies, and our internal expertise, to support the entirety of the Department of the Air Force (DAF) Air and Space Forces.

As the DAF and DoW move forward with Joint All Domain Operations and human-autonomous technologies as key focus areas, RHW is a key partner in tackling these challenges. RHW addresses the warfighters needs to be adaptable, efficiently trained, provided decision-quality information, and have technologies that augment their abilities. There are common threads traversing training, preparedness, and operations; RHW's expertise and capabilities address these connective tissues.

This past year our team supported multiple exercises, organized and presented at conferences, conducted studies, developed software, hit multiple scientific and technical transition milestones, and demoed novel capabilities to the warfighter... Congrats to the entire RHW Team!

I look forward to sharing future Warfighter Interactions and Readiness Division successes with you! ☆

Mr. Dean "Bear" Berry, Division Chief, 711 HPW/RHW

RHW **ORGANIZATION**



DIVISION CHIEF
Mr. Dean Berry



DIRECTOR OF OPERATIONS
Maj Taylor Paige, Ph.D.



SENIOR ENLISTED LEADER
MSgt Jacob Scott

BRANCH LEADERSHIP

TECHNICAL ORGANIZATION

PRODUCT LINES (PL)



READINESS
Mr. David Malek



WARFIGHTER-MACHINE INTEGRATION
Dr. Sarah Bibyk



INTELLIGENCE ANALYTICS & SENSEMAKING
Dr. Jerred Holt

CORE TECHNICAL COMPETENCIES (CTC)



HUMAN LEARNING & COGNITION
Dr. Brian Simpson

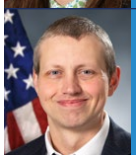


WARFIGHTER INTERFACES & TEAMING
Dr. Mark Draper

CORE RESEARCH AREAS (CRA)



DIGITAL MODELS OF COGNITION
Dr. Megan Morris



LEARNING & OPERATIONAL TRAINING
Dr. Christopher Stevens



DISTRIBUTED TEAMING & COMMUNICATION
Dr. Mike Tolston



HUMAN-MACHINE INTERACTIONS
Mr. Peter Venero

BRANCHES

RHWE



COGNITION AND MODELING
- **Ms. Jennifer Brown** (BC)
- **Maj Ramone Williams** (DO)

RHWO



OPERATIONAL LEARNING SCIENCES
- **Dr. John Camp** (BC)
- **Maj Michael Corpuz** (DO)

RHWT



COLLABORATIVE TECHNOLOGIES
- **Dr. Terry Stanard** (BC)
- **Lt Col Kip Johnson** (DO)

RHWI



WARFIGHTER INTERFACES
- **Dr. William "Paul" Murdock** (BC)
- **Maj Amanda Thiel** (DO)

BC = BRANCH CHIEF | DO = DIRECTOR OF OPERATIONS

NOVEMBER 2026

READINESS
PRODUCT LINE



Mr. David Malek
Readiness PL Lead,
711 HPW/RHW

Welcome to the 2026 edition of *Fights On!* and to the Readiness Product Line update for this year. During a time of renewed focus on operational relevance, the Readiness Product Line remains dedicated to a single, critical mission: ensuring our Airmen and Guardians possess the decisive cognitive and performance advantages needed to dominate any future conflict. Our commitment to a decisive edge extends deep into advanced modeling and simulation and our focus remains squarely on the warfighter—developing and extending the technologies that enhance proficiency for current and future operational challenges. The work highlighted in the following pages reflects our commitment to forging the world's most formidable fighting force.

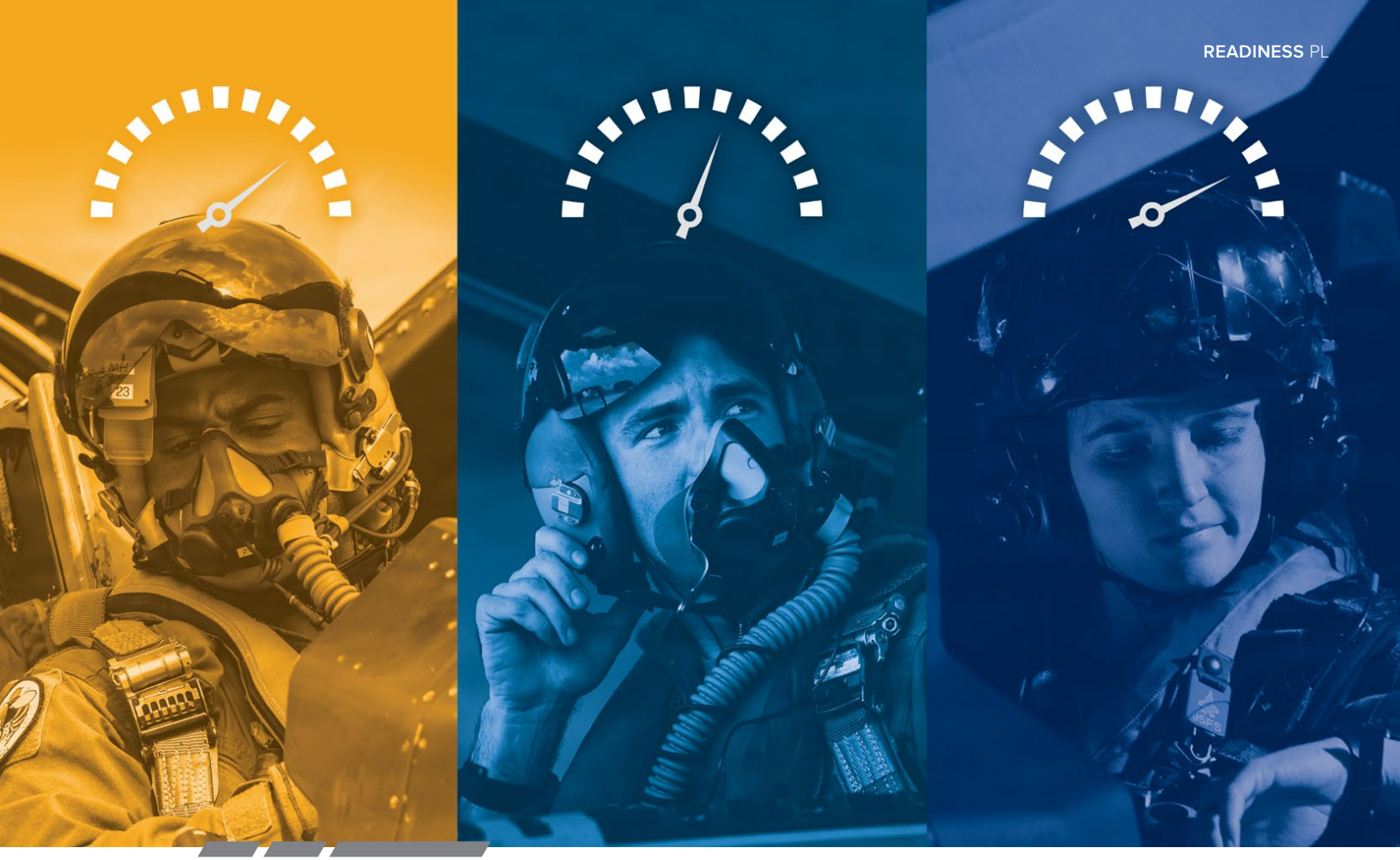
This past year has been one of tremendous progress, driven by the evolving needs of the warfighter. The **Fighter Integration Training Research (FITR)** effort, detailed by Dr. Beth M. Hartzler and Mr. Scott "Carp" Carpenter, connects 5th-generation (F-22) and 4th-generation (F-16) virtual simulators to create a more integrated air-to-air warfighting environment so pilots and Air Battle Managers/Weapons Directors (ABM/WDs) experience integrated mission briefing, execution, and debrief while we at AFRL conduct our research into training methods and tools. Over the past 12 months the team conducted two fighter-integration and one fighter-simulation training research weeks, collecting data from four F-22 pilots, 14 F-16 pilots, and four ABM/WDs—slightly more than half of them active duty, the remainder Air National Guard. Observed changes include improvements in weapons allocation, threat management, and adherence to operational guidelines, along with a significant decrease in self-reported stress. A new questionnaire on Collaborative Combat Aircraft (CCA) found that both pilots and ABM/WDs expect CCAs to be most beneficial as additional weapons, tools to extend sensor range, and a means of adding mass to confuse adversary targeting. The training concern of greatest agreement was teaching operators to manage the added cognitive workload of controlling multiple aircraft. Looking to 2027, the testbed will receive updated operational flight programs, high-fidelity F-22 full-mission trainers, CCA and mixed-reality capabilities, F-35 devices on the fighter-integration network, and a data rich, competency-driven debrief dashboard.

PL OVERVIEW

The Readiness Product Line develops and extends technologies and tools for improving cognitive effectiveness, performance, and proficiency of Airmen and Guardians in current and future operational mission contexts.

The Tailored Regimens for Adaptive Instructional Needs (TRAIN) program described by Dr. Tiffany Jastrzembski-Myers applies artificial intelligence to convert training activity into mission-relevant readiness signals. TRAIN is domain-agnostic: cognitive models, machine learning, and large language models combine to assess, adapt, and optimize learning wherever performance data exist, with applications spanning linguist training, bystander-intervention training, pilot training, medical training, and human-machine teaming. Results to date include test-score improvements of up to 81 percent, a 30 percent increase in learner motivation, and a 75 percent improvement in converting fragmented data into actionable training guidance.

Continued on next page



Graphic by Mr. Will Graver (SierTek, Ltd.)

The system builds on two AFRL patents and has progressed to Technology Readiness Level 4–6. Capabilities are in use or under evaluation at the Defense Language Institute, AF/A1Z, and the Netherlands Aerospace Centre.

Our foundational work for the **Surrogate Training Family of Systems (STFoS)** – some are returning to calling the program REFORGE – continues to mature Proficiency-Based Training (PBT) tools while providing the 7th Fighter Training Squadron (7th FTS) a state-of-the-art training capability. In their article, Dr. Summer Rebensky and Mr. Jordan Haralson describe delivery of a multi-platform training device (MPTD), a first step toward "agnostic cockpits," or simulator cockpits that can be rapidly reconfigured, in this case, from a T-38A to an F-22. In this iteration, the MPTD is configured to include T-38A emergency-procedure simulation for the 7th FTS in which switches relevant to emergency procedures are interactable. Using Mission Essential Competency workshop outputs and artificial intelligence-enabled generation reduced 7th FTS mission-qualification and instructor-pilot upgrade syllabus updates from months to weeks. AFRL is now working with ACC/A3 to demonstrate a potential enterprise syllabus-management tool that can generate graduate evaluations, digital gradesheets, and Ready Aircrew Program (RAP) requirement lists. Smart gradesheets on pilot electronic flight bags fused with currency data in a readiness dashboard let the squadron rapidly determine who is combat-mission-ready. The 7th FTS will transition to our REFORGE-inspired simulation ecosystem as its exclusive training system beginning in FY27, with these capabilities slated to scale to other operational squadrons and Air Force readiness dashboards.

Finally, work with the **Simulator Common Architecture Requirements and Standards (SCARS)** program continues to discover the art of the possible and provide recommendations to the SCARS acquisition program. Mr. Brad Pfefferle and Mr. Cameron Roudebush report that the Joint Simulation Environment–On-Premises Equipment (JSE OPE)—a high-fidelity simulation server purpose-built to run Global Reusable Interface Doman (GRID) services, GPU-intensive applications, and mission-training software—has been configured within the AFRL SCARS testbed. In parallel, Rapid Common Tool Configuration research is delivering Ansible-based automation that creates SCARS-compliant virtual machines and configures applications such as TaskView, Performance Evaluation and Tracking System (PETS), live, virtual, and constructive (LVC) Network Control Suite (LNCS), and the Advanced Simulation Combat Operations Trainer (ASCOT) from the SCARS Library. Containerized software-testing environments and centralized data-repository pipelines for high-resolution imagery and terrain models further accelerate authorization-to-operate pathways and give affiliates a continuously updated, high-fidelity simulation environment.

Each of these initiatives, from PBT ecosystems and integrated 4th/5th-generation simulation, to common architecture standards and adaptive AI, represents a critical thread in the fabric of total readiness. The following articles provide a deeper look into the work our team and our partners have accomplished this year. We are proud of our achievements and energized to face the challenges of 2027 as we continue to deliver the training tools and methods that forge the world's most formidable fighting force. ☆

Mr. David Malek, Readiness Product Line Lead, 711 HPW/RHW

Fighter Integration Training Research (FITR)



One component of the Readiness Product Line is the Fighter Integration Training Research (FITR) effort which focuses on improving training technologies and resources for DAF Airmen and Guardians. "Integration" refers to the connection between our 5th (F-22) and 4th (F-16) generation flight simulators, allowing pilots and air battle managers (ABMs) to experience mutual mission briefing, execution, and debrief situations seldom available outside of combat operations.

Participants attending our training research weeks (TRWs) typically complete six simulated missions, operating against constructive entities, per day for 2 to 4 days. In addition to the objective performance data these scenarios generate, we collect subjective responses regarding stress and workload during the scenarios, recent training and combat experience, and objectives for the TRW. These measures enable us to better contextualize performance data and improve the foundational understanding of workload and readiness to integrate additional capabilities.

In the past 12 months we have conducted three TRWs, collecting data from four F-22 pilots, 14 F-16 pilots, and four ABMs. Slightly more than half of all TRW attendees were active duty while the remainder represented Air National Guard units. Despite differences in scenario composition and mission profiles, there have been some generalizable findings such as pilots demonstrating improved weapons allocation, threat management, and adherence to operational guidelines. Additionally, some measures of self-reported workload during the simulated missions indicated a significant decrease in stress level. These data will serve as an important basis for comparison as the testbed continues to mature through updated operational flight programs (OFPs), the integration of new capabilities, and changes to simulator fidelity and display technology.

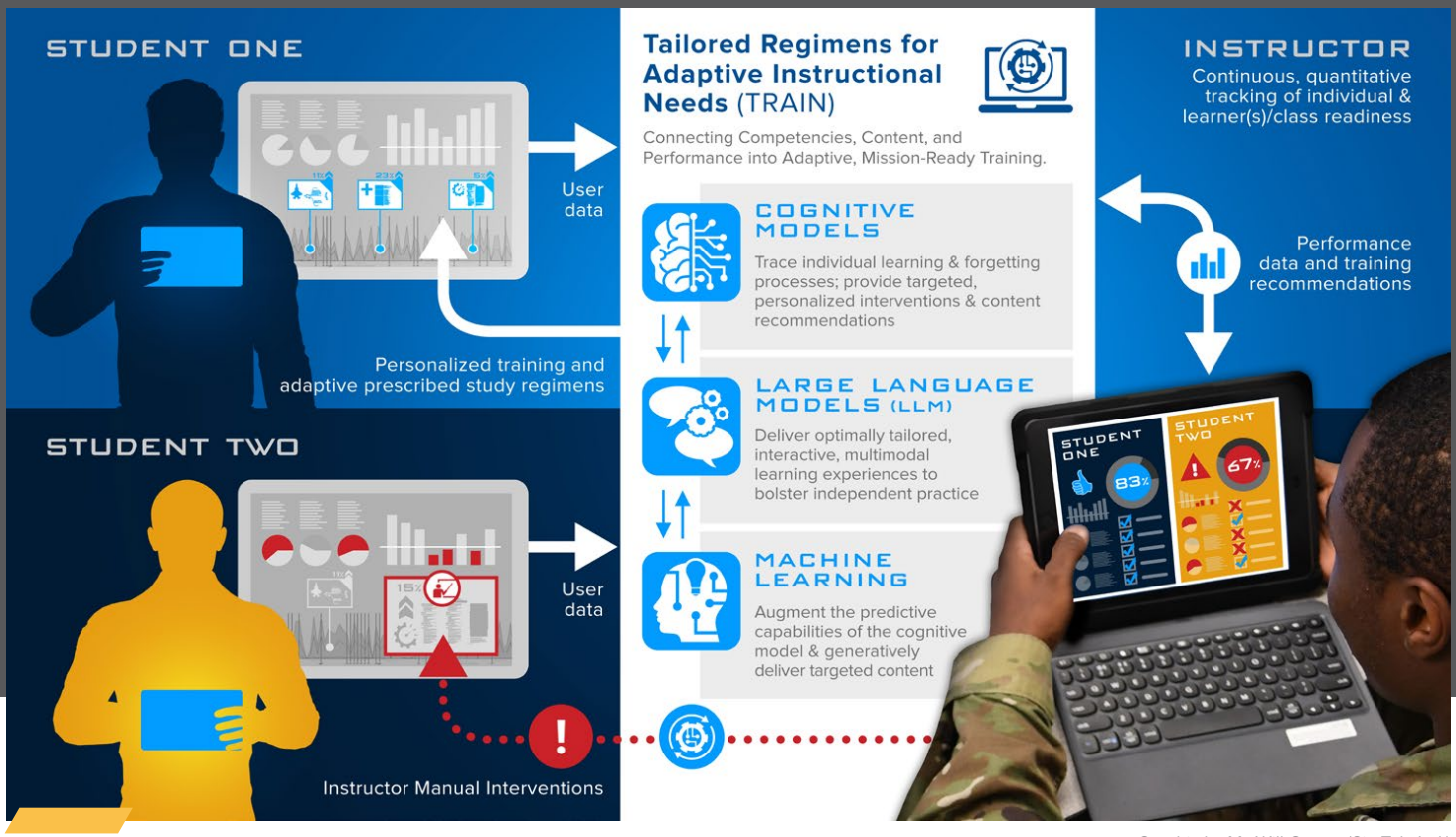
The FITR effort is unique in that it offers airmen both short- and long-term benefits. The immediate advantage comes in the form of providing additional training opportunities for TRW attendees. When asked about their objectives for participating in a TRW, pilots and ABMs most commonly report a desire to maintain mission readiness, receive feedback on performance, and learn about the capabilities of other aircraft, with most ultimately reporting their goals were satisfied. More important though are insights gained for implementation of novel capabilities, methods of evaluating and predicting performance, and the efficacy of novel training technologies.

For example, this year we implemented a new measure asking pilots and ABMs about their expectations for the role of collaborative combat aircraft (CCA). Also known as "loyal wingmen," CCAs will be semi-autonomous, with oversight and direction from fighter pilots in the area and are intended to augment existing forces or assume risks inappropriate for manned aircraft. Our questionnaire asked about anticipated uses for CCAs during combat, primary training concerns, and potential solutions to problems encountered during operation. Both pilots and ABMs anticipate CCAs will be most beneficial as a source of additional weapons, tools to extend sensor range, and adding mass to confuse adversary targeting. Likewise, pilots and ABMs agreed the training area of greatest concern would be teaching operators to effectively manage the added cognitive workload of controlling multiple aircraft.

Moving into 2027, our research will expand through changes to our testbed and debrief capabilities. Updated OFPs have been installed in the F-22 and F-16 simulators and are currently awaiting pilot validation. Moreover, we will replace the existing medium-fidelity F-22 cockpits with high-fidelity full-mission trainers and thereby increase the immersive quality and operational validity of the experience. Other planned additions to the testbed include the integration of CCA capabilities, development of mixed reality simulators, and finalization of F-35 simulators that will join the FI network. Finally, we will evaluate the training utility of new debrief capabilities by implementing a data-driven dashboard summarizing details such as shot performance, kill ratio, and expanding shot logs to reflect modern weapons. These changes represent critical steps in addressing the DAF's evolving capabilities and training needs. ☆

Dr. Beth M. Hartzler, Research Psychologist, 711HPW/RHWOW

Mr. Scott "Carp" Carpenter, Fighter SME, RCG



Graphic by Mr. Will Graver (SierTek, Ltd.)

TRAIN: Applied AI for Adaptive, Mission-Focused Readiness

The Tailored Regimens for Adaptive Instructional Needs (TRAIN) research program is focused on improving readiness through applied artificial intelligence (AI). At its core, TRAIN is a domain-agnostic, adaptive learning and assessment infrastructure designed to convert training activities into mission-relevant readiness signals, rapidly build targeted skills, and optimize performance acquisition and sustainment. Rather than relying only on course completion or subjective judgment, TRAIN helps translate fragmented or low-fidelity training data into actionable intelligence that can support more precise, personalized, and effective training decisions.

What makes TRAIN especially valuable is that the capability is not tied to a single schoolhouse, mission set, or learner population. The underlying approach integrates cognitive models, machine learning, and large language models to create a generalizable ecosystem for human performance optimization. In practical terms, the domain may change, but the core capability does not: TRAIN can assess, adapt, and optimize learning and readiness wherever relevant performance data exists. Current and emerging applications span linguist training, bystander intervention training for the total force, pilot training, medical training, and additional mission areas tied to human-machine teaming and readiness modernization.

Results to date show strong promise. TRAIN-enabled adaptive AI has demonstrated improvements in test scores up to 81 percent, increased learner motivation by 30 percent, and improved the conversion of fragmented data into actionable readiness intelligence by 75 percent. The system builds on two AFRL patents and has progressed to Technology Readiness Level (TRL) 4–6, with additional growth expected as more operationally relevant data becomes available. Capabilities are already in use or under test at the Defense Language Institute (DLI), AF/A1Z, the Netherlands Aerospace Centre (NLR), and the technology is being implemented as a cornerstone of the DLI ecosystem.

TRAIN has also produced concrete validation evidence in aviation-related applications. TRAIN researchers established cognitive assessment pipelines predictive of pilot success, demonstrated specific cognitive-skill validity for pilot outcomes, and developed subject matter expert (SME)-constrained AI that translated pilot performance data into trackable, high-fidelity profiles. Through collaborative work with the Netherlands, high-fidelity AI tracking of empirical pilot skills achieved r = 0.86, demonstrating a foundation to next assess pilot training optimization through targeted remediation or augmentation.

Just as important, TRAIN possesses modular capabilities. Its measures, models, assessment pipelines, and adaptive training functions can be tailored and integrated into new operational settings without rebuilding the entire ecosystem from scratch. That modularity is what makes the program well suited for expansion. We are actively seeking new domains and application areas where TRAIN's adaptive, proficiency-based capabilities can improve performance, reduce training costs, increase throughput, and strengthen mission readiness. ☆

Dr. Tiffany Jastrzembski-Myers, Senior Cognitive Scientist, 711 HPW/RHWE

Scaling Readiness: Surrogate Training Family of Systems

Readiness Focused Tracking Systems

Airmen readiness has only become more critical—and remains challenging to track in real time while providing prescriptive insights for training optimization. The AFRL and Air Combat Command (ACC) continue to advance the Surrogate Training Family of Systems (STFoS) effort. Through a close partnership with the 7th Fighter Training Squadron (7th FTS), AFRL's Proficiency-Based Training (PBT) methods have been applied to deliver a readiness first training ecosystem for the Air Force.

Proficiency Enabled Training Platforms

In 2025, AFRL delivered a multi-platform training device capable of quickly transitioning from a T-38A configuration to an F-22. This capability enabled lightweight training options and facilitated the assessment of extended reality (XR) performance in flight training. This year, the 7th FTS received a custom T-38A simulator from AFRL designed specifically for emergency procedure training. As a "goldilocks" style solution, only switches relevant to emergency procedures are interactable, allowing for a low cost virtual reality training capability.

The simulator is supported by a suite of technologies from PBT methods that expedite scenario startup, collect necessary metadata for long-term training tracking, and automatically assess emergency procedure performance. These capabilities not only provide a new simulation training capability for the 7th FTS but also establish foundational technologies for weaponeering procedure tracking and assessment, along with data requirements for upcoming fighter platforms.

Revamping Training Development and Certification

With the rise of artificial intelligence (AI) and the growing need for timely, relevant training tracking capabilities, AFRL has been exploring AI applications across the full training pipeline. Using outputs from Mission Essential Competency (MEC) Workshops, AFRL evaluated structured Retrieval Augmented Generation (RAG) workflows to populate syllabus sections within Headquarters Air Force (HAF) standard templates. AI-enabled workflows reduced updates to the 7th FTS mission qualification and instructor pilot upgrade syllabi from months to weeks.

Since then, AFRL has been working with Headhquarters AF A3 to develop an enterprise level syllabus management tool. In addition to producing syllabi in standard formats, these workflows can generate other essential outputs—such as graduate evaluations, digital gradesheets, and complete Ready Aircrew Program (RAP) training

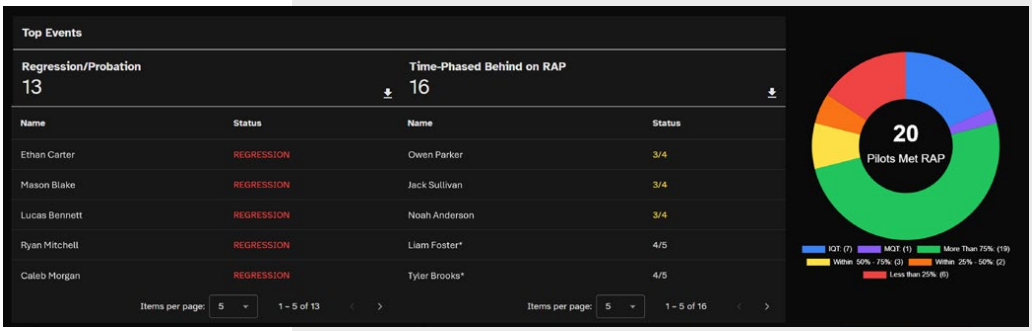
Continued on next page





Photos (above) by Dr. Summer Rebensky and the 7th Fighter Training Squadron. Screen capture (below) by Dr. Summer Rebensky

Developed dashboards synthesize training, Squadron Aviation Resource Management (SARM), stan eval, gradesheets, Training Accomplishment Record (TAR) sheets, and Artemis exports for a one-stop squadron tracking shop.



STFoS Ecosystem



Adaptive Mission Training Technologies

- XR immersive training devices with data enabled capabilities for training and automatic performance measurement

Readiness Focused Tracking Systems

- AI-enabled syllabus generation of proficiency-based advancement centric training programs & automatic reporting for sim certification
- Mission Essential Competency (MEC) focused grade sheets centered on CAF standard debrief

Synthesis of Critical Information

- Integration of squadron shop information (e.g., training, SARM, Stan eval, flying hour tracking program) for real-time squadron status updates

Prescriptive Training and Analytics

- Highlighting and prescribing individual and squadron level needs to inform scheduling needs for proactive and readiness-based training pipeline

Graphic by Mr. Will Graver (SierTek, Ltd.)

requirement lists—helping reduce cross organization friction and support broader Air Force needs. As integration with the 7th FTS continues to improve, these technologies will be leveraged in new ways to enable more efficient execution of squadron processes and reporting requirements. For example, the STFoS System could directly synthesize, validate, and generate content for the End of Year RAP Closeout report, significantly reducing the workload for unit training staff members and shortening the overall reporting process.

Those digital RAP training requirements support the Air Force's Simcert process, in which each weapon system simulator is evaluated against the training task lists (TTLs) to provide a comprehensive breakdown of which training tasks can be effectively accomplished in the simulator. AFRL has collaborated with the 29th Training Evaluation Squadron (TES) to produce expedited simulator evaluation forms, real time dashboards, and AI-enabled report generation. In 2026, AFRL supported multiple evaluations, including an acceptability test report for the Joint Synthetic Environment, which was completed in a matter of days using automated reporting tools. These tools are currently being refined for adoption by the 29 TES to enhance the Simcert process moving forward.

Training Tracking and Prescriptive Analytics

Integration of squadron shop information [e.g., training, Squadron Aviation Resource Management (SARM), Stan/Eval, flying hour tracking] for real-time squadron status updates is often manual. Identifying training gaps remains difficult when operational and training flights occur daily—shifting the target while ensuring all training and currency requirements are met. AFRL has delivered integrated dashboard solutions that highlight and prescribe individual and squadron level needs, informing scheduling decisions for a proactive and readiness-focused training pipeline.

Smart gradesheets, available on pilot Electronic Flight Bags (EFBs), follow the Combat Air Force standard debrief flow: grading only maneuvers accomplished; requiring comments for competencies below the required proficiency level; and flagging regressions or unaccomplished maneuvers within a training block. Each of these features helps instructors quickly identify next sortie priorities.

Gradesheet data and pilot currency data combine in a readiness dashboard that enables the squadron to rapidly determine who is combat mission ready and who has met required training benchmarks. The 7th FTS will transition to the STFoS ecosystem as its exclusive training system beginning in FY27. The capabilities described here are slated to scale to other operational squadrons as well as provide HAF level AF readiness dashboards. These solutions, developed by AFRL, bring training needs to the forefront to ensure the most effective investment of AF training dollars and improve Airman readiness. ☆

Dr. Summer Rebensky, Principal Scientist and Portfolio Lead, Aptima, Inc.
Mr. Jordan Haralson, Program Manager, Aptima, Inc.

AFRL Simulator Common Architecture Requirements and Standards (SCARS)

AFRL Simulator Common Architecture Requirements and Standards (SCARS) started in 2017. The research edge of SCARS works to discover the art of the possible. The team has provided guidance and conducted research across Air Force solutions to provide recommendations and standardizations, increasing awareness of existing solutions across a variety of platforms.

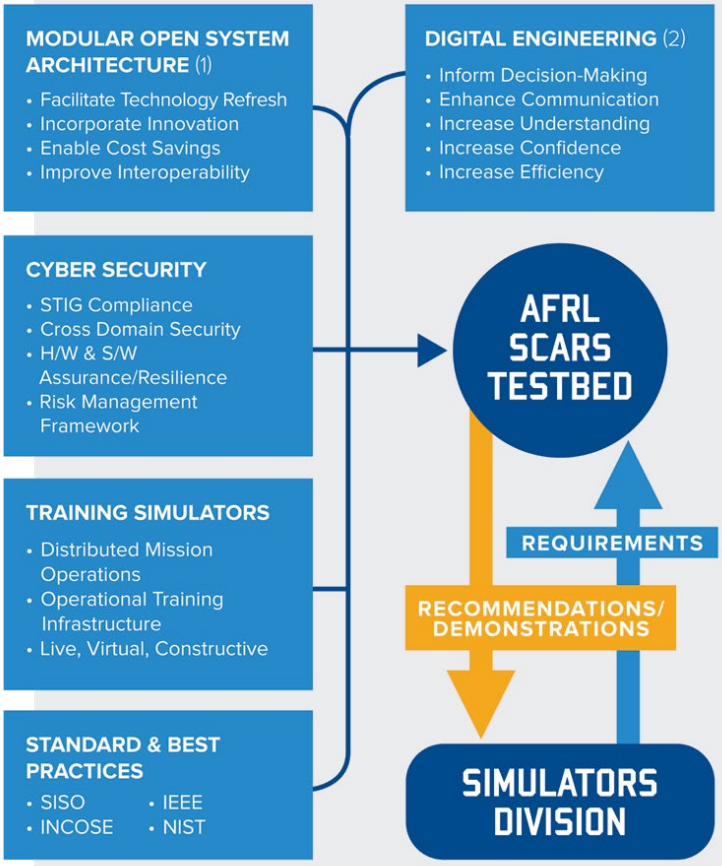
JSE OPE Experiments

The Joint Simulation Environment – On Premises Equipment (JSE OPE) is a high fidelity simulation server purpose built to run Global Reusable Interface Doman (GRID) services, GPU intensive applications, and mission training software within the AFRL SCARS ecosystem. It provides a localized, high performance environment where training centers can host virtual machines, execute complex sensor models, and synchronize truth data across distributed training sites. The JSE OPE has recently been configured within the SCARS testbed, enabling teams to run TaskView, Advanced Simulation Combat Operations Trainer (ASCOT), GRID services, and other high demand workloads for realistic and scalable simulation analysis.

Current efforts on the JSE OPE focus on verifying resource allocations, testing virtual machine (VM) configurations, and evaluating how well different setups support varying volumes of training activity. This work includes using Virtual Local Area Network (VLAN) isolated sites, simulated Wide Area Network (WAN) delays, and GRID service correlation tests to understand performance limits and ensure accuracy across distributed sites. These evaluations directly strengthen training reliability—allowing the AFRL SCARS testbed to identify optimal configurations, prevent bottlenecks, and ensure consistent sensor and environment modeling across all participating simulators. This testing enables SCARS to benchmark architectures, validate GPU dependent applications prior to fielding, and mature deployment automation across environments. The impact is clear: more predictable performance, stronger interoperability, and higher readiness for upcoming training events that depend on precise, high throughput simulation capabilities.

Rapid Common Tool Configuration Research

The Rapid Common Tool Configuration effort is directly improving the usefulness and accessibility of standardized training tools across Air Force training centers. By developing Ansible based automation that creates SCARS compliant VMs and configures applications such as TaskView, Performance Evaluation and Tracking System (PETS), LVC (Live-Virtual-Constructive) Network Control Suite (LNCS), and ASCOT from the SCARS Library—including tools already carrying Certificate to Field (CTF)—AFRL is significantly reducing the setup burden on operators and site administrators. Training centers gain the ability to rapidly field consistent, properly configured environments without relying on specialized technical knowledge, ensuring that new or updated capabilities can be used much sooner and with higher reliability.



References: Defense Acquisition Guidebook Chapter 3
DoD Digital Engineering Strategy, June 2018

This work enhances **standardization, repeatability, and confidence** across OPE sites. By validating these automation workflows within the SCARS Testbed and ensuring they meet SCARS standards and OPE constraints, AFRL is delivering a trusted, reproducible method for deploying common training applications at multiple classification levels. The result is far less manual configuration, fewer opportunities for misconfiguration, and faster availability of mission relevant tools for instructors, operators, and analysts. As training centers continue adopting these scripted deployment methods, overall readiness is strengthened through more reliable tool usage, reduced downtime, and streamlined integration of future SCARS Library resources.

To maximize the SCARS OPE's value across the broader enterprise, current testing focuses on integrating foundational capability pipelines that benefit both government stakeholders and program affiliates. By validating containerized software testing environments, the OPE ensures developers and site configurators have immediate access to secure, pre-hardened software containers—accelerating Authority to Operate (ATO) pathways and standardizing cybersecurity compliance across sites. Concurrently, testing centralized data repository pipelines establishes a unified mechanism to deliver high-resolution imagery and terrain model data directly to all participating affiliates. Modernizing legacy data distribution methods with this shared baseline provides affiliates with a continuously updated, high-fidelity simulation environment in a highly cost-effective manner for the government. ☆

Mr. Brad Pfefferle, Principal Software Engineer, Aptima, Inc.
Mr. Cameron Roudebush, Senior Software Engineer, Aptima, Inc.



Dr. Sarah Bibyk
Warfighter-Machine Integration
PL Lead, 711 HPW/RHW

WARFIGHTER-MACHINE INTEGRATION

PRODUCT LINE

It has been another productive year for the programs under the Warfighter Machine Integration Product Line. With the retirement of Dr. Deirdre Mahle last year, it has been an honor and a privilege to step into the role of Product Line Lead for this remarkable portfolio. In this overview article I will highlight just a sample of the major accomplishments achieved by the teams this past year.

The **Machine Operator Blending for Integrated Uncrewed Systems (MOBIUS)** team has continued to press with their research and development of adaptable, scalable, and intuitive user-focused human-machine interfaces, coming off of the success of several flight tests performed by Australian partners from the Defence Science & Technology Group (DSTG), the Royal Australian Air Force (RAAF) and Boeing Defence Australia (BDA) to demonstrate ground and airborne Command and Control (C2) of Collaborative Combat Aircraft (CCA). The team, led by Kristen 'Cuda' Barrera, has been working with the Air Force CCA Program of Record to accelerate the A-GRA (Autonomy-Government Reference Architecture) compliance of their technical solutions. The team also successfully established a Cooperative Research and Development Agreement (CRADA) to better facilitate the collaborative development of Human-Machine Interface (HMI) solutions for CCA management with industry partners.

Continued on next page





Photos (left) by Kristen "Cuda" Barrera and (right) the Royal Australian Air Force

The **Joint All Domain Persistent Adaptive Collaborative Technologies (JADPACT)** team, led by Dr. Elizabeth Frost, has continued spearheading their ongoing research and development effort, the DASH (Decision Advantage Sprint for Human-Machine Teaming) events. This year the events culminated in the first ever MASH (Multi-Decision Advantage Sprint for Human-Machine Teaming) event in May, as detailed more fully in this year's Summer issue of Fight's On! (Issue 78). The MASH event brought together the Department of the Air Force's Advanced Battle Management System (ABMS) Cross-Functional Team; the 805th Combat Training Squadron (CTS); the JADPACT, Human Language Technologies (HLT), and Team Metrics teams from AFRL's Human Effectiveness Directorate (RH); and the Coordinating Austere Nodes through Virtualization and Analysis of Streams (CANVAS) team from AFRL's Information and Spectrum Warfare Directorate (RF) Directorate. As part of the 2-week software sprint and data collection event, software solutions informed by the Transformational Model of Decision Advantage (TM-DA) for battle management were solicited from vendors who were able to work directly with the AFRL teams and operators during the sprint to develop and integrate their microservice solutions into the JADPACT HMI. These experiments continue to be vital for assessing how machine support can reduce decision time and improve decision quality dramatically for air battle managers working in complex operational environments.

This upcoming year will see many more critical milestones for these programs. I am excited for what the future holds for this brilliant team of scientists, engineers, and developers as they pursue the cutting-edge HMI solutions in battle management and C2 for our warfighters. ☆

Dr. Sarah Bibyk, Warfighter-Machine Integration PL Lead, 711 HPW/RHW

PL OVERVIEW

Identify and mature operator centric interfaces and human-machine teaming that increase Airmen and Guardians' combat capabilities through integrated solutions that develop synergies, maximize battlespace interoperability, maintain situation awareness, and increase combat power and decision dominance while decreasing cognitive workload.









Photos (top three) by Deb Henley and (bottom) by Airman 1st Class Jennifer Nesbitt

INTELLIGENCE, ANALYTICS & SENSEMAKING
PRODUCT LINE



Dr. Jerred Holt
Intelligence, Analytics & Sensemaking
PL Lead, 711 HPW/RHW

The Intelligence, Analysis and Sensemaking (IAS) Product Line continues to accelerate multiple science and technology (S&T) efforts and has already demonstrated solid operational mission impact. With a strong baseline in traditional Intelligence, Surveillance, and Reconnaissance (ISR) process augmentation through the Legion program, our focus has been on increasing involvement in the Space and Cognitive Warfare domains as well as building better interconnectivity between research and development (R&D) systems.

PL OVERVIEW

Captures the full spectrum of data collection and exploitation, including novel intelligence planning/ orchestration, enhanced analysis through advanced human-machine teaming, and rapid sensemaking through warfighter-centric analytic tools, AI/ML/LLM based agents, and visualization technologies in support of decision superiority. Leverages maturing technologies in augmented cognition, trust, and decision-making to develop Cognitive Warfare tools and influence applications.

The primary goal of IAS is to provide decision superiority to the AF and Joint intel community through tools, systems and process that augment the operator's ability to make timely, effective decisions in an increasingly organic and distributed battlespace. The diversity of expertise, responsibility, and communication options present in the intel process presents an unparalleled challenge to our ability to execute complex kill webs with persistent target custody.

The **Legion** program is a suite of services that aim to optimize the process of developing and executing intelligence requirements. It provides multi-role decision support through advanced modeling and simulation, databasing and artificial intelligence and machine learning (AI/ML) backends, coupled with empirically driven Human-Machine Interfaces (HMI). After a successful demonstrations, Legion has seen increased partnerships both within the AFRL and our operational partners. These include deployments to US Pacific Command (PACOM), US Indo-Pacific Command (INDOPACOM), US European Command (EUCOM), as well as tech integrations with AFRL efforts such as APEX Gear, SAINT, and Resolute Sentry.

Two key programs were stood up within the IAS PL that included **Best of Breed Configurable Adaptive Teammate (BOBCAT)** and **MORRIGAN**. Both programs leverage years of AFRL investments in foundational R&D to deliver capabilities that address critical AF demand signals. BOBCAT, as the name suggests, is an adaptive teammate architecture built on high fidelity speech interpretation/synthesis and bespoke large language models (LLM). It aims to provide a configurable framework that can support multiple operator roles and tasks. MORRIGAN focuses on strategic measures of cognitive warfare. The team will be building on a pedigree of RH Information Warfare research to provide a predictive modeling capability to support Indications & Warnings of enemy behavior. ☆

Dr. Jerred Holt, Intelligence, Analytics & Sensemaking PL Lead
and Senior Research Psychologist, 711 HPW/RHW





Dr. Glenn Gunzelmann

Modeling & Simulation Lead,
Warfighter Interactions & Readiness
Division, 711 HPW/RHW

MODELING AND SIMULATION INTEGRATION LAB (MSIL)

Winning the wars of the future requires urgency and speed in the operationalization of new technologies. Our innovation must outpace our adversaries. However, it is vital that we recognize that this speed in discovering, developing, and delivering new technologies is necessary but not sufficient for maintaining our advantage. Technology must be paired with human operators to maximize our effectiveness.

The Modeling and Simulation Integration Lab (MSIL) is the Department of the Air Force's (DAF's) premier science & technology (S&T) environment for warfighter-centric modeling, simulation, and analysis infrastructure and experimentation to optimize human-technology integration, transition, test, and training. Our mission is to develop and deliver game-changing warfighter-in-the-loop digital modeling, simulation, and analysis capabilities that equip Airmen & Guardians to fly, fight, & win now and in the future.

The goal of the MSIL is to develop a prototype operational training and test infrastructure (OTTI) that meets the requirements for digital capabilities supporting high-end advanced training, test, and tactics (HEAT3) for the DAF. It consists of a physical environment supporting multiple levels of security, technology integration, and experimentation with humans in the loop, coupled with a modeling & simulation (M&S) architecture – the Flexible Architecture for Synthetic Training and Test (FASTT). The architecture provides the tools and specifications for creating the ecosystem that enables the DAF to train and test how we fight and the digital environment that allows our technology to mature and transition at the speed of need. The MSIL infrastructure, including FASTT, supports applied and advanced technology research for human system integration, human machine teaming, trust, training, and decision making across operational domains. In addition, it provides a venue for evaluating and demonstrating technologies in a warfighter-centric environment. In addition to providing a lower cost, digital option for these research activities, MSIL also provides a front door to the DAF's digital thread through alignment with the Joint Simulation Environment (JSE), a critical M&S backbone that connects S&T to OTTI. The benefits are improved integration between human performance and warfighting capabilities in an environment that speeds the transition of those capabilities into operations.

We are building collaborations across the AFRL, Air Force Materiel Command (AFMC), and the DAF to promulgate our infrastructure concepts in pursuit of shared specifications and standards that promote interoperability and common data foundations across the acquisition and training ecosystem. In addition, we are building cross-service collaborations to promote joint all-domain modeling, simulation, and analysis that preserves our advantage into the future. ☆

Dr. Glenn Gunzelmann, Principal Research Psychologist, 711 HPW/RHW



Photos by Mr. Will Graver (SierTek, Ltd)



Dr. Brian Simpson
Human Learning & Cognition CTC
Lead, 711 HPW/RHW

HUMAN LEARNING & COGNITION

CORE TECHNICAL COMPETENCY (CTC)

Comprehensive operational readiness remains one of the greatest challenges facing the US Air and Space Forces today. Success in the future fight will require solutions and training that evolve as quickly as the threats we will face, and as technological parity has become a reality for our adversaries, our ability to prevail will be determined by the complete, and dominant, readiness of the *Human Warfighter*. The Human Learning and Cognition Core Technical Competency (HLC CTC) provides the tools and capabilities to support warfighter readiness through a program of research focused on developing and integrating computational models of human information processing and advanced training technologies into a comprehensive readiness ecosystem.

Future warfighters will need technologies and training solutions that can adapt as quickly as the adversaries and challenges they confront. This requires the creation of an advanced, extensible, multi-domain training ecosystem capable of replicating the operational complexities anticipated in future Joint All-Domain mission environments – that is, we must be able to *train as we fight*. Critical to this vision is the ability to create personalized training tools that can support rapid skill acquisition and retention as well as ensure warfighter proficiency and availability. We must develop advanced training architectures and data management systems that allow for the collection of contextualized, purposeful data to address metrics identified as critical in all phases of learning and yield analytics for developmental training intervention. We must also create the technologies that support human/Artificial Intelligence (AI) teams that can rapidly and effectively learn from, and with, each other in real time for adaptive learning that may be required during critical mission operations.

Train as we fight also means that we must effectively recreate *all* aspects of a mission space – most critically, the human element – in order to reach complete operational fidelity, and integrate that human element into a larger synthetic

Continued on next page

> **Our Airmen must fly, fix, and train with purpose, realism, and discipline to ensure we are ready any time our nation calls."**
>
> **— General Kenneth S. Wilsbach**
> Chief of Staff, United States Air Force

CTC OVERVIEW

Promote more lethal Air and Space Forces through research on human multisensory perception, learning, information processing, and action. Maximize mission effectiveness through personalized, proficiency-based readiness for multi-capable Airmen and Guardians in joint all-domain operations, enable teams of humans and machines to adapt and learn together in real time in training and operational settings, and advance considerations of human performance in system development and operational planning with digital models of perception, cognition, and action.



Graphic by Mr. Will Graver (SierTek, Ltd)

Human Learning & Cognition CTC

Digital Models of Cognition CRA

Learning & Operational Training CRA

modeling and simulation training architecture. The implementation of high-fidelity representations of human teammates and adversaries depends on the development of scalable models of perceptual and cognitive processes that mediate decision-making. These processes are impacted by the many stressors present in operational environments, including cognitive load and fatigue; physical factors such as operating at altitude, in noise, constrained by personal protective equipment; uncertainty; and the challenges associated with team coordination and, importantly, the realities of skill decay.

The HLC CTC addresses the above challenges by organizing our research portfolio into two focused Core Research Areas (CRAs), each of which has identified a set of goals in response to demand signals from DoD leadership:

1. **Digital Models of Cognition**: This CRA addresses the DoD's emphasis on a digital transformation, enabling this team to rapidly and effectively support elite operational training and assess weapons systems with human models in place in order to predict performance and mission success and inform proactive and reactive mitigation strategies.
2. **Learning and Operational Training**: This CRA represents the AFRL's historical expertise in research for elite operational training supporting Warfighters in air, space, and cyberspace, addressing leadership's demand for persistent readiness that can only be achieved through large-scale, Joint All-Domain operational training exercises at a frequent cadence.

The HLC CTC research portfolio addresses the most challenging, and time-critical, issues identified by USAF and USSF leadership. The following sections provide a more detailed description of our efforts and demonstrate how we will continue to deliver capabilities to the warfighter in order to position the DoD to be the strongest fighting force in the world. ☆

Dr. Brian Simpson, Human Learning & Cognition CTC Lead, 711 HPW/RHW



Dr. Megan Morris
Digital Models of Cognition
CRA Lead, 711 HPW/RHWE

Digital Models of Cognition
CRA

Holistic Models for Decision-Making
LOE

Multi-Sensory Human System Modeling and Evaluation
LOE

Information Mastery in Cognitive Warfare
LOE

DIGITAL MODELS OF COGNITION

CORE RESEARCH AREA



Strategic competition with our adversaries demands our warfighters to be agile, swift, resilient, and decisive. To prepare our warfighters and to ensure unbeatable decision-making during missions, we must develop informative, real-time models that simulate cognitive and social influence processes at different abstraction levels and generalize across domains with realistic internal and environmental stressor effects. The Digital Models of Cognition (DMC) Core Research Area (CRA) addresses this need by developing and advancing computational and mathematical modeling capabilities that represent human perception, cognitive-social information processing, and behavior in warfighter decision-making in operational contexts. We transition knowledge products, software, and tools to the warfighter and other customers to ensure readiness and enhance performance. We have three technical challenge areas (Lines of Effort) that drive our CRA work:

CRA OVERVIEW

Emphasize research to identify computational and mathematical mechanisms that represent human perception, information processing, and behavior. Integrate models to reflect the role of internal and external factors that modulate performance efficiency and effectiveness. Develop holistic models that capture a host of cognitive and social mechanisms and support quantitative understanding and prediction of decision-making effectiveness across domains and at different levels of abstraction for improved systems engineering, operational planning, wargaming, and training.

1. The first challenge area is **Holistic Models for Decision-Making (HMDM)** which focuses on addressing warfighter cognitive overload by developing knowledge products, models, and software that can detect and predict cognitive workload and inform mitigation strategies to enhance decision-making performance.
2. The second is **Multi-Sensory Human System Modeling and Evaluation (MSHSME)**, a fully externally funded LOE that supports requests from customers focused on evaluating the physical characteristics of technologies and human systems to address evaluation gaps and enhance human performance and safety in complex operational environments.
3. The third challenge area is **Information Mastery in Cognitive Warfare (IMCW)** which focuses on developing models, analytic methods, and tradecraft that enables operators to improve Information-Related Capability for the desired cognitive effects, social influence, and mastery of the Information Environment.

The CRA is made up of a unique, technically diverse team, and further embraces a multidisciplinary approach by collaborating with several teams across the AFRL and other services, industry, and academia. Over the last year we have worked to focus our research on specific technical challenges and domains within the LOEs to efficiently and effectively use our manpower to address critical warfighter needs. In the following articles, you will learn more about our LOEs and we will highlight some specific efforts for a deeper dive. ☆

Dr. Megan Morris, Digital Models of Cognition CRA Lead, 711 HPW/RHWE



Graphic by Mr. Will Graver (SierTek, Ltd)

HOLISTIC MODELS FOR DECISION-MAKING

LINE OF EFFORT

The Holistic Models for Decision-Making (HMDM) Line of Effort (LOE) has had an exciting year pursuing research that aims to understand warfighter decision-making capabilities in complex and cognitively demanding operational tasks. Our team continues to explore the full range of human decision-making abilities, especially under high cognitive load and particularly when compounded by the addition of physiological and external stressors. Through this research, our team pursues which technology and training solutions will be most effective to augment or supplement human decision-making capabilities and mitigate potential failures in decision-making during operations.

The multi-disciplinary LOE team addresses research questions across a variety of topics, including visual processing, auditory processing, learning, fatigue, and attention during multi-tasking. A brief summary of these projects follows. The project *Environment, Noise Demands, Utilization of Cognitive Resources, and Extended Mission Time for Mobility Crew* (ENDURE-MC) completed a successful workload and acoustic noise data collection with a Combat Systems Officer (CSO) while in flight, the results of which are currently being fully analyzed, with a planned follow-up study in an in-lab simulator. The *Grimoire* project (formerly AE4AI), which is a cross-LOE effort pursuing solutions for complex analysis tools, has expanded into large language model solutions to rapidly extract reliable insights from open-ended operator data, and it is currently being deployed within human behavioral studies to analyze subjective pilot feedback, while transitioning into a scalable microservice platform for repeatable, advanced statistical modeling. The Hybrid-Cognitive Models project is another cross-LOE research effort and focuses on developing adaptive, multi-model approaches that better predict a learner's knowledge acquisition, in collaboration with the Ohio State University.

In the articles that follow, we will take a deeper dive into a few more specific findings. In *Impact of Night Vision Goggle Use on Sleep, Cognition, and Performance*, our researchers investigate differences in long term use of white vs. green phosphor NVGs on a variety of tasks. In *Modeling the Command of Multiple Autonomous Collaborative Platforms*, we discuss the development of models to capture operator decision-making while managing ACPs. Finally, in *A Physiocognitive Model of Modafinil's Effects on Fatigue* our researchers demonstrate a novel physiocognitive model that captures effects of modafinil on task performance during sleep deprivation. ☆

Dr. Sarah Bibyk, Former HMDM LOE Lead, 711 HPW/RHWEO

Program Leads

LOE LEAD
Current: Dr. Megan Morris
Former: Dr. Sarah Bibyk

PROGRAM MANAGER
1st Lt Isabella Colizzi



Figure 1. Example of blooming in green phosphor NVGs



Figure 2. Example of scintillation in white phosphor NVGs

Photo (top) by Sgt. Eric Pahon, 82nd Combat Aviation Brigade and (bottom) by Sharon Ellis and Frederick Meyer, 711 HPW/RHWE

For more information on this effort, please refer to the following paper:

Curley, T., Poole, E., Ellis, S., Meyer, F., Fox, E., Dennis, B., Robinson, H., Sharma, K. & Holderby, K. (2026). Impacts of long-term night vision goggle use on fatigue and cognitive performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* https://doi.org/10.1177/10711813261475165

Impact of Night Vision Goggle Use on Sleep, Cognition, and Performance

Night Vision Goggles (NVGs) are a critical component of night-time military special operations, allowing operators to detect objects normally too dark for the human visual system to perceive. The increase in operations tempo requires operators to utilize NVGs over long periods of time, across multiple days. The original NVG eyepiece display image screen, similar to the original green cathode ray tube (CRT) computer monitor display screen, has been changed from green, Figure 1, to white, Figure 2, by changing the display phosphor.

The change from green to white phosphor included performance improvements in image intensifier tube technology and has drawn accolades from the user community. How the visual differences between NVGs with different phosphor colors affects operator performance over longer duration missions is relatively unknown. Naval Medical Research Unit-Dayton (NAMRU-D), collaborating with the Holistic Models for Decision-Making (HMDM) Visual Perception Modeling Team, examined performance for sustained vigilance, visual acuity and working memory through the Psychomotor Vigilance Test (PVT)- Reaction Time (RT), PVT- Lapse Probability (LP), Dynamic Visual Acuity Test and N-Back Test. The tests consisted of a series of tasks, performed across three eight-hour sessions separated by a week, with participants alternating between unaided (No- NVGs), green, and white phosphor NVGs. The experiment consisted of a series of tasks, performed through three eight-hour sessions, separated by a week with participants alternating between white or green phosphor NVGs. Results indicate performance on cognitive tasks is similar between the two NVG types in the short term but better for participants using the green phosphor NVGs over long term. The results of the study provide novel and specific estimates of visual acuity, working memory, and vigilance for different NVG types over sustained periods of time. These findings offer critical insights to inform performance predictions that optimize mission execution, enhance safety, and model aspects for insight with future advanced technologies. ☆

Dr. Taylor Curley, Research Psychologist, 711 HPW/RHWEM
CDR Erica Poole, PhD, NAMRU-D, Wright-Patterson, AFB
Ms. Sharon Ellis, Senior Research Psychologist, 711 HPW/RHWEO
Mr. Frederick Meyer, Electronics Engineer, 711 HPW/RHWEO
Dr. Elizabeth Fox, Research Psychologist, 711 HPW/RHWEM
Ms. Bethany Dennis, CTR, Leidos, Inc.
Ms. Heather Robinson, CTR, Leidos, Inc.
Dr. Kharananda Sharma, CTR, Leidos, Inc.
Dr. Katherine Holderby, Oak Ridge Institute for Science and Education, Oak Ridge, TN



Graphic by Mr. Will Graver (SierTek, Ltd)

Modeling the Command of Multiple Autonomous Collaborative Platforms

During flight, pilots and aircrew are required to engage in quick and frequent decision-making while maintaining high situational awareness, leading to increased cognitive load and fatigue. These difficulties are compounded by the rapid development of Autonomous Collaborative Platforms (ACPs), or drone wingmen, which will require aircrew to simultaneously manage multiple autonomous aircraft. Thus, it is imperative to understand how the command of multiple ACPs affects decision-making and to identify strategies that win the fight without overloading pilots.

Researchers in the Holistic Models for Decision-Making Line of Effort are using machine learning and computational techniques driven by principles of cognitive performance to examine these issues. First, the Task Activity Tracker (TAT) translates raw telemetry data from the Advanced Framework for Simulation, Integration and Modeling (AFSIM) into dynamic, real-time visualizations of pilot performance when using Pilot-Vehicle Interfaces (PVIs). By visually tracing the "breadcrumbs" of a pilot's actions through a menu tree in real-time, TAT instantly highlights interface bottlenecks, inefficient command routing, and navigational errors. The ultimate goal of TAT is to pinpoint impediments to pilot performance, such as cognitive overload, and highlight inefficiencies in ACP pilot vehicle interfaces (PVIs) that limit efficient human-machine teaming.

Second, the Models of Autonomous Control and Heuristics (MACH) framework uses a complex reinforcement learning (RL) architecture to investigate “optimal” policies in commanding multiple ACPs. MACH simulates pilot behaviors in commanding multiple ACPs, as well as their own ship, during air-to-air combat against a variable number of enemy aircraft (2-8). The RL agent’s behaviors are shaped by rewards for positive outcomes (e.g., eliminating an enemy aircraft), penalties for negative outcomes (e.g., losing an ACP), and cognitively plausible constraints, such as command execution delays (2-5 seconds). Over millions of trials, the agent learns the best ways to maximize their rewards until they consistently earn the highest rewards possible – the “optimal policy”. After training the model, researchers can replay episodes and find out “why” agents made certain decisions using explainable artificial intelligence (XAI) methods, such as SHAP (SHapley Additive exPlanations) values and integrated gradients, that attribute environmental variables to the actions that they influence. MACH then uses AFRL’s Google Vertex AI machine learning platform to examine model performance, analyze tactics that emerge from the policy, and synthesize the results into a detailed after-action report.

Overall, TAT and MACH, combined with human behavioral data collected at the 711th Human Performance Wing, provide state-of-the-art analytical techniques to examine complex pilot behaviors when controlling multiple ACPs using dedicated PVIs. The insights from these models will help optimize ACP interfaces and accelerate tactics development for the future fight. ☆

Dr. Taylor Curley, Research Psychologist, 711 HPW/RHWEM
Dr. Adam Fenton, Research Psychologist, 711 HPW/RHWEM
Dr. Chris Myers, Senior Cognitive Scientist, 711 HPW/RHWEM

A Physiocognitive Model of Modafinil's Effects on Fatigue

Cognitive fatigue is a critical factor in long-duration missions, increasing safety risk and degrading mission performance. Stimulants are a common means to mitigate fatigue, prolonging wakefulness when sleep is not an option. In order to effectively utilize stimulants in long-duration missions to enhance operational effectiveness, we must be able to model their effects on performance over time given the amount of extended wakefulness, time of day, and warfighter physical characteristics (i.e., weight). A paucity of models exist for the stimulant modafinil, which is less addictive, has fewer cardiovascular issues, and minimal effects on subsequent sleep as compared to caffeine and dextroamphetamine. The Holistic Models for Decision-Making Physiocognitive Modeling team proposed an effort to the 711 HPW/USAFSAM Studies and Analysis Program and was awarded. This effort aimed to develop a physiocognitive model of modafinil's effects on fatigue that combines both physiological and cognitive model architectures to accurately imitate these stimulants with behavior-specific predictions based on cognition. Specifically, we examined the generalizability of an existing caffeine physiocognitive model with an archival dataset that included active duty military performing the Psychomotor Vigilance Test (PVT) during a 36-hour total sleep deprivation period where modafinil had been administered, and modified the existing model to accurately reflect modafinil's behavior. Additionally, we used this new modafinil physiocognitive model to inform changes to an abstracted biomathematical fatigue model that is commonly used in fatigue risk management tools.

The original analyses for the archival dataset partitioned PVT performance into high and low performer groups. Using these metrics, we compared the existing caffeine physiocognitive model predictions to the PVT reaction time behavior in the modafinil dataset. We found that the model had good fits for median reaction time and lapses (reaction times > 500 ms) but poor fits for false starts (reaction times < 100 ms) suggesting the model would need to be modified to accurately account for modafinil effects. We then reconstructed an existing minimal physiologically based pharmacokinetic (PBPK) model of modafinil from the literature and used this PBPK model to make changes to the cognitive model. These changes resulted in good fits for high performer data and marginal fits for low performer data. We then used the new modafinil physiocognitive model to modulate a biomathematical fatigue model that is commonly used in fatigue risk management tools. Again, we saw a similar pattern with good fits for high performer data and marginal fits for low performer data due to a delayed effect of modafinil in the observed data compared to what was predicted for low performers. However, rank ordering of predictions for lower performers was accurate, suggesting overall acceptable model performance. We compared the modafinil biomathematical model to the baseline biomathematical model to compare predictive accuracy. Overall, the modafinil SAFTE model performed better than the baseline model especially for high performers but had mixed results for low performers.

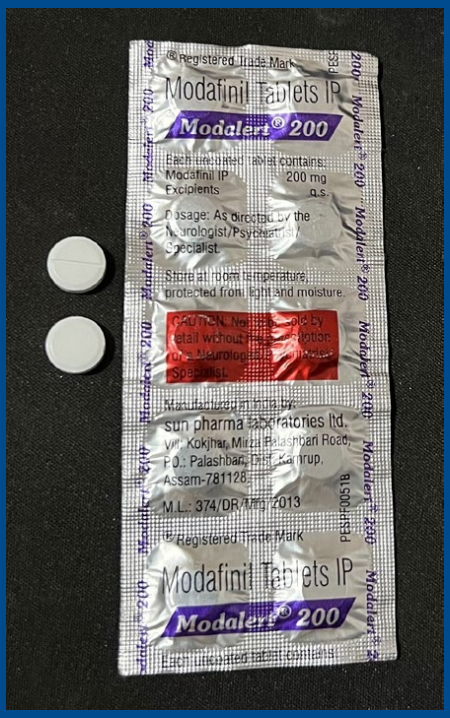


Image by WikiLinuz,
Source: https://commons.wikimedia.org/wiki/File:Modafinil_200mg.jpg

This new modeling capability has several applications. For example, we can generate what-if scenarios for modafinil ingestion to help develop more useful and effective dosing guidelines during long duration missions. We can model real-time fatigue assessment and prediction updates to crew members based on their actual modafinil ingestion. We can also use the model for after-action review to retrospectively assess cognitive fatigue if modafinil was ingested during the mission – this can be useful for examining the utility of dosing regimens and potentially help mishap investigations involving the use of modafinil. Future work will be needed to enhance predictive accuracy for low performers and test the modeling capability against other modafinil datasets and with real-time data. ☆

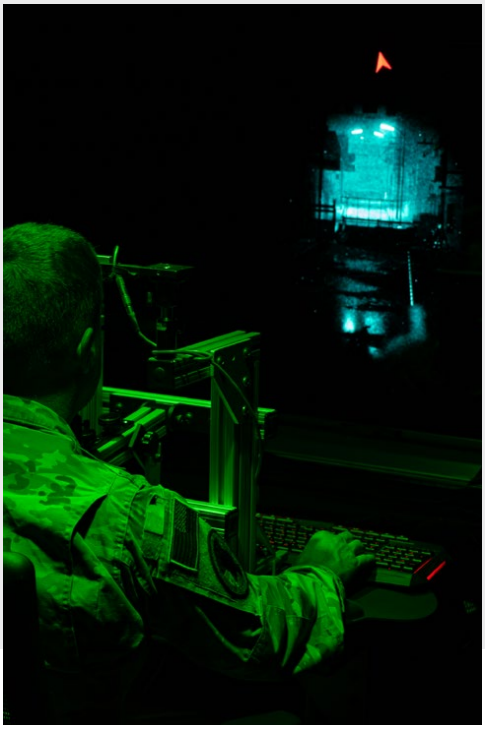
For more information on this effort, please refer to the following papers:

Halverson, T., Morris, M. B., Myers, C. W., & Sweeney, L. (in press). Physiocognitive modeling: Explaining the effects of modafinil on fatigue. *Proceedings of the 24th International Conference on Cognitive Modelling.*

Morris, M. B., Sweeney, L. M., Halverson, T., Davis, I., Myers, C. W., Curley, T., & Caldwell, J. L. (2026). Integrating modafinil effects into a biomathematical model of fatigue. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* https://doi.org/10.1177/107118132614752

Dr. Megan Morris, Digital Models of Cognition CRA Lead, 711 HPW/RHWE
Dr. Christopher Myers, Sr. Cognitive Scientist, 711 HPW/RHWEM
Dr. Taylor Curley, Research Psychologist, 711 HPW/RHWEM
Dr. Tim Halverson, CTR, Aptima, Inc.
Dr. Lisa Sweeney, CTR, AV Inc.
Dr. Lynn Caldwell, CTR, Leidos
Mr. Ian Davis, CTR, BAE Systems, Inc.





Photos (left) courtesy of SpaceX, (middle) SSgt Isaiah Soliz, and (right) by Rick Eldridge, 711 HPW CAG

MULTI-SENSORY HUMAN SYSTEM MODELING AND EVALUATION

LINE OF EFFORT

The Multi-Sensory Human System Modeling and Evaluation (MSHSME) Line of Effort (LOE) is made up of five primary research areas that support technology and human system evaluations for performance and safety, as well as provide datasets to the Holistic Models for Decision-Making LOE for cognitive load software development, validation, and verification. The first area focuses on F-35 noise data and impact assessment on Warfighter personnel with the F-35 Joint Program Office as the primary stakeholder, international partners/FMS participants as secondary stakeholders, and international collaborators. The second area focuses on noise exposure/impacts of missiles and space vehicles on the surrounding environment with primary stakeholders, Japanese Self Defense Forces, USSF Space Launch Deltas 30 and 45, and Air Force Civil Engineering Center, as well as several other stakeholders and collaborators in the Federal Aviation Administration, NASA, and the Department of Transportation. The third area focuses on developing acoustic sensor algorithms and systems for counter-drone operations. The fourth area focuses on evaluation of technology and perceptual factors that describe/affect night vision goggle performance with stakeholders in AFLCMC/WNY. The fifth area focuses on evaluation of hearing protection and communication device and complex noise exposures on Warfighter health and performance with stakeholders across the Department of Defense. ☆

Dr. Megan Morris, MSHSME LOE Lead, 711 HPW/RHWE
Ms. Hilary Gallagher, MSHSME PI, 711 HPW/RHWEM
Dr. Alan Wall, MSHSME PI, 711 HPW/RHWEO
Dr. Frank Mobley, MSHSME PI, 711 HPW/RHWEO
Ms. Sharon Ellis, MSHSME PI, 711 HPW/RHWEO

Program Leads

LOE LEAD
Dr. Megan Morris

PROGRAM MANAGER
Ms. Kayelin Tiggs



Graphic by Mr. Will Graver (SierTek, Ltd)

INFORMATION MASTERY IN COGNITIVE WARFARE

LINE OF EFFORT

The Information Mastery in Cognitive Warfare (IMCW) Line of Effort (LOE) is conducting research and developing solutions to support Information Operations missions. This diverse group is working to provide improved capabilities for operators within the information environment through research and development by delivering practical tools and techniques that will enhance readiness and mission effectiveness. Outputs from this LOE will feed into our new MORRIGAN project, an artificial intelligence (AI)-powered tool suite that will assist in influence operations analysis, planning, and assessment.

One project within this LOE is the development of a simulation of the operational Information Environment in collaboration with the 39th Information Operations Squadron. This project is intended to create a representation of the critical aspects of the information environment for training, exercises, and wargaming. The simulation will allow teams to practice and refine their methods in a controlled environment that can model various information-based scenarios, generating appropriate responses by simulations of target audiences, providing a valuable resource for preparing for the complexities of actual operations.

In addition to the simulation, the IMCW LOE is creating tools to improve the assessment of Information Operations in collaboration with the Joint Military Information Support Operations (MISO) WebOps Center at the Special Operations Command (SOCOM). These tools will leverage the Grimoire toolkit and large language models, and are being designed for use during the planning, execution, and assessment phases of an operation. The objective is to give commanders and planners a clearer understanding of the effects of their actions. By providing more accurate and timely feedback, these tools will help personnel evaluate the impact of their work and make more informed decisions.

A third effort focuses on bolstering the cognitive security of airmen. The LOE is conducting research on the cognitive and social science aspects of social engineering to develop training programs, educational resources, and models that enhance the force's ability to recognize and resist adversary influence operations, such as disinformation and propaganda with new models generating weaponized hyper-personalized content at speeds unimaginable just a few years ago. The project seeks to provide software and digital tools and methods to reduce the vulnerability of military personnel to hostile information activities, thereby strengthening the overall resilience of the force. The following article will focus on one of these cognitive security efforts. ☆

Dr. Eric Thompson, IMCW LOE Lead, 711 HPW/RHWE

Program Leads

LOE LEAD
Dr. Eric Thompson

PROGRAM MANAGER
Capt Addison Shableski



Screen capture by Dr. Fairul Mohd-Zaid, 711 HPW/RHWEM

Autonomous OSINT and IO Planning Tool

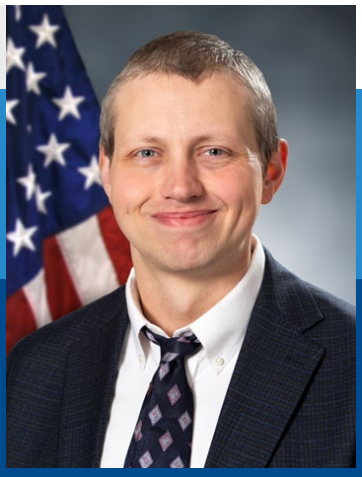
The modern Information Environment presents a complex, fast-moving battlespace where decision advantage relies on rapid sense-making. Presently, operators and planners are at risk of being overwhelmed by the sheer volume and velocity of Open-Source Intelligence (OSINT). Accurate and swift decision-making is critical to countering adversarial information tactics, but manual data processing creates an operational bottleneck that threatens mission success. As a newly developed proof of concept, the Autonomous OSINT & Information Operations (IO) Planning Tool aims to reduce that gap by demonstrating how an analytical assistant can automate intelligence synthesis to directly feed into an operator's decision process.

Structured analytical frameworks are essential for various military applications, particularly when evaluating civil considerations and mapping operational terrain. However, building these baseline analyses often requires exhaustive research that may not be readily available to every analyst during high-tempo operations. Building on previous work under the Information Mastery for Cognitive Warfare LOE, this research focuses on enhancing automated analysis for the Joint Planning Process (JPP). The project provides a bench-level analytical service tailored for 14F Joint IO Planners and Joint Intelligence (J2) and Joint Operations (J3) directorates. By leveraging Agentic Artificial Intelligence (AI) and Large Language Models (LLMs), the prototype is capable of engineering precise prompts derived from the Commander's Intent and trusted OSINT source to generate comprehensive Political, Military, Economic, Social, Information, Infrastructure, Physical, and Time (PMESII-PT) and Areas, Structures, Capabilities, Organizations, People, and Events (ASCOPE) assessments.

Specifically, the Autonomous OSINT & IO Planning Tool demonstrates the capability to streamline data ingestion, effectively condensing days of manual research into minutes, and expediting findings within complex information environments. This successfully proves how AI can enhance the evaluation of adversary vulnerabilities and target audience dynamics. To ensure future operational reliability, the architecture incorporates a strict "zero-hallucination" framework that enforces explicit citations, linking every generated intelligence assessment back to its root domain and source report to build immediate trust with J2 teams. Furthermore, it validates the ability to transform raw data into dynamic geospatial map overlays equipped with color-coded PMESII rosettes and custom Military Standard (MIL-STD) symbology, offering a glimpse into a future capability for instant, actionable situational awareness.

While this work directly supports Department of Defense (DoD) priorities by showcasing advanced capabilities designed for the tactical edge, the tool currently exists as a successful prototype. Designed for flexible deployment, the system has proven it can operate on secure, air-gapped networks using local models, or interface with web-hosted AI frameworks. In the interim, the tool feeds into broader product line efforts such as MORRIGAN which is an overarching Intelligent Analytics and Sensemaking system that supports 14F operations planning by applying social science theories to analyze aggregated demographic, survey, and publicly available data. Additionally, this work supports the Simulated Non-kinetic Agent Response Environment (SNARE) project, which is backed by Congressional Interest Item (CII) funding for AI applications in Cognitive Warfare. By successfully automating robust intelligence framework generation, this proof of concept lays the vital groundwork for future integration across multiple joint systems. ☆

Dr. Fairul Mohd-Zaid, Mathematical Statistician, 711 HPW/RHWEM



Dr. Christopher Stevens
Learning & Operational Training
CRA Lead, 711 HPW/RHWO

Learning and Operational Training
CRA

Co-Learning for Adaptive Human-Machine Teams
LOE

Warfighter Learning Technologies
LOE

Gaming Research Integration for Learning Laboratory ®
GRILL

LEARNING & OPERATIONAL TRAINING

CORE RESEARCH AREA

The wars of tomorrow will be fought by teams consisting of humans and autonomous agents. These human–machine teams must understand each other, learn new concepts flexibly, adapt to uncertain, rapidly evolving combat situations, and maintain decision superiority. Training technologies and methodologies must reflect this operational reality. The Learning and Operational Training (LOT) Core Research Area (CRA) advances learning and cognitive readiness within future operational environments. Specifically, our research supports Operational Training and Test Infrastructure (OTTI) ecosystems designed to maximize mission effectiveness, minimize training costs, personalize requirements, and optimize human–machine teaming.

CRA OVERVIEW

Establish an ecosystem that maximizes mission effectiveness while minimizing costs by matching technologies to learning and performance needs; supporting high resolution human and system measurement and quantitative, proficiency-centric readiness assessment and prediction at the individual and team levels. Explore how to enable human and machine co-learning to support mutual adaptation and understanding in human-machine teams.

Warfighters must be prepared to fight and win across increasingly complex, multidomain operations. This environment introduces significant scientific and technical challenges for operational training:

- **Performance Metrics**: What quantifiable indicators define elite performance in individuals, teams, and heterogeneous human–artificial intelligence (AI) units?
- **Real-Time Adaptation**: How can instructional systems adapt dynamically in real time to maximize warfighter lethality and readiness?
- **Capability Integration**: How do we most effectively train the unique combinations of skills demanded by emerging operational capabilities?

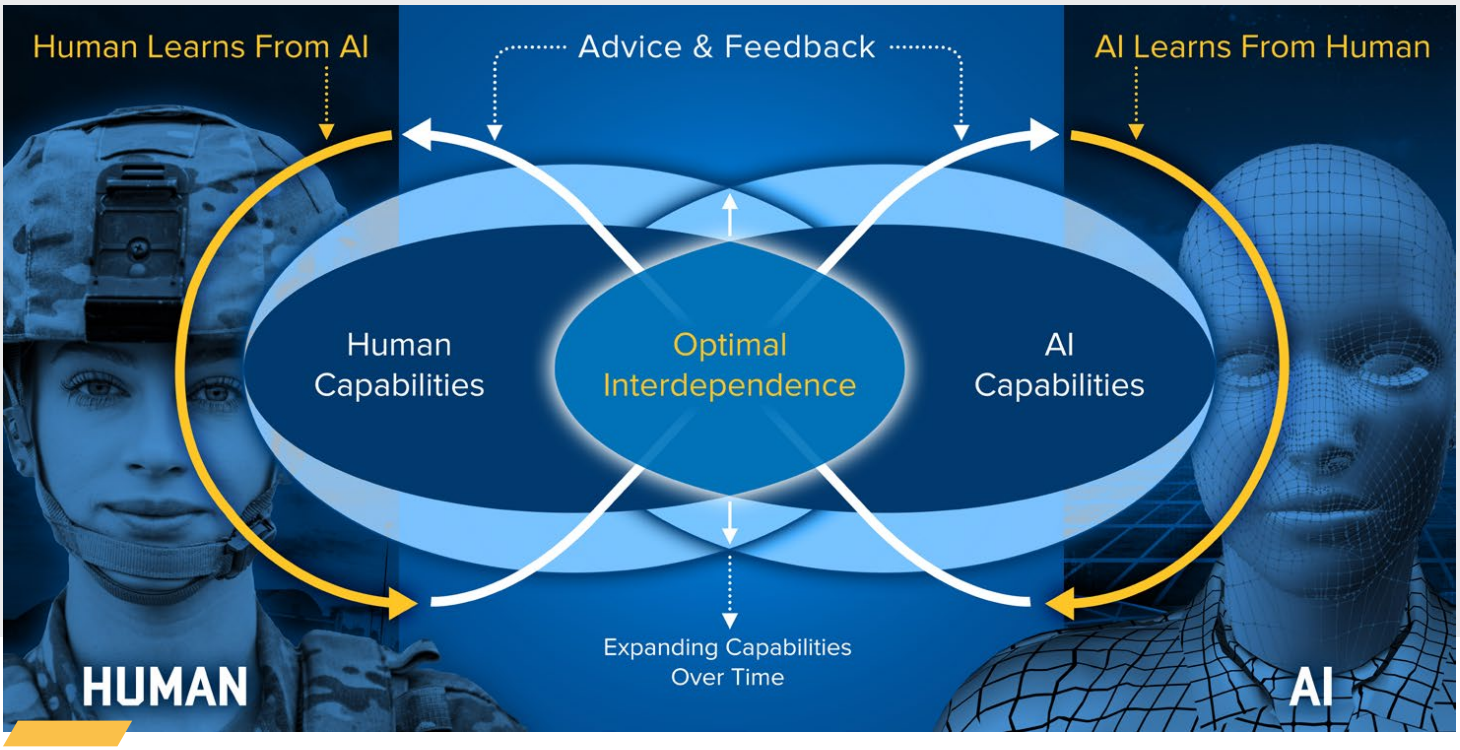
To address these challenges, the LOT CRA leverages cognitive and learning sciences to deliver analytics and interventions for robust proficiency acquisition and assessment. Current efforts include model-based assessments tracking proficiency in undergraduate and operational flight training, multi-crew communication, and intelligence analysis. These analytics rapidly identify operational gaps and prescribe tailored training regimens.

Furthermore, LOT researchers are developing novel tools leveraging generative AI to create adaptive training scenarios dynamically while mitigating risks such as hallucinations. A key initiative within this portfolio is the Just-In-Time Tactical Tutor (JINTTACT), which monitors proficiency across distributed human and synthetic teammates to target critical competency shortfalls. Complementing this work, the team integrates proficiency models with large language model (LLM) pipelines to power intelligent training ranges capable of automated After-Action Review (AAR) and adaptive syllabus orchestration during warfighter exercises, including Large Force Exercises (LFEs).

Future operational training and test environments will not only tailor training delivery to each warfighter's individual trajectory, but also dynamically reconfigure instructional conditions as the operational landscape shifts. Harnessing AI's operational potential is vital to achieving and sustaining decision superiority. ☆

Dr. Christopher Stevens
Learning and Operational Training CRA Lead and Senior Research Psychologist, 711 HPW/RHWO

The Co-learning Cycle



Graphic by Capt Raymart de Asis, 711 HPW/RHWOH

CO-LEARNING FOR ADAPTIVE HUMAN-MACHINE TEAMS

LINE OF EFFORT

The conundrum: artificial intelligence (AI) provides the warfighter with unprecedented operational advantages while also introducing new and consequential vulnerabilities. That is, AI compresses the observe, orient, decide, and act (OODA) loop to improve the speed and scale of decision support. However, AI systems can be fragile or imperfect partners when lacking the capacity to learn from human teammates. At present, AI cannot independently correct factual errors nor incorporate essential human judgment as events unfold. Without a mechanism for ongoing learning, AI may operate on static or incomplete information that can result in high-regret decisions, obscure newly emerging opportunities, and remain vulnerable to adversary manipulation.

One solution: Co-learning technologies address these limitations through a real-time, reciprocal process in which human and machine teammates learn from one another to improve decision-making during inherently uncertain and dynamic warfare. Here, machines provide computational speed, scale, data fusion, and analytic reasoning, while humans contribute qualitative judgment, context, and interpretation of intelligence and other relevant information. Although humans can readily learn from both human and machine teammates, the central challenge is enabling machines to learn from human-provided knowledge at operationally relevant time scales. Advancing co-learning capabilities is essential to achieving effective human-machine teaming and ensuring that AI-supported decisions remain adaptive, informed, and responsive to changing conditions.

Program Leads

LOE LEAD
Dr. Lorraine Borghetti

PROGRAM MANAGER
2d Lt Joseph Bowman

Advancements: Recent co-learning innovations include the **AI-enabled Model of the Opponent**, an intent-prediction engine, and **Oko**, a sophisticated space-based simulation environment for testing co-learning concepts and AI models; these technologies were featured in the *Fight's On!* Summer 2026 issue. In the current issue, the Co-learning team presents several additional co-learning technologies: the **human-interactive machine learning architecture (HIMLA)**; **Co-learning Refined AI (CRAI)**, an adaptive long short-term memory (LSTM) model; an **analysis of large language model behavior in the Ultimatum Game**; and the **Multi-domain Adaptive Text Traversal for Open-Source Collection & Knowledge (MATTOCK)**, a government-tailored web-scraping tool. ☆

Dr. Lorraine Borghetti, Senior Research Scientist, Co-Learning for Adaptive Human-Machine Teams LOE Lead, 711 HPW/RHWOH

2d Lt Joseph Bowman, Co-learning Program Manager, 711 HPW/RHWOH

Novel Co-learning Technology Integrates Deep Learning with Human-guided Bayesian Inference

Closing the gap between human insight and machine-speed decision support requires artificial intelligence (AI) systems that can learn from their human teammates as conditions change. To meet that challenge, the Co-learning team partnered with the University of Dayton Research Institute (UDRI), the Air Force Institute of Technology (AFIT), and the AFRL Space Warfare Directorate to develop a real-time machine-learning solution that combines deep learning and Bayesian inference.

Model development focused on a space domain awareness (SDA) scenario in cislunar space. SDA in this environment is challenging because objects may change states, move beyond sensor coverage, and follow nonrepetitive trajectories. Limited and imperfect sensors can produce sparse, noisy, or incomplete data. Moreover, small errors in initial-state estimates may cause substantial trajectory divergence over time. Consequently, numerical estimates of object location and behavior have limited utility in isolation and require interpretation alongside operational context, uncertainty, and likely intent.

Operator assessments of likely intent can help constrain this problem by defining or prioritizing plausible hypotheses. Simulated trajectories and sensor observations can then be used to estimate Bayesian likelihoods associated with each hypothesis. For example, if a satellite is hypothesized to be approaching a cislunar disposal area or surveilling a lunar asset, the model can assess how well observed behavior aligns with that intent. This approach reduces an otherwise vast set of possible trajectories to a more manageable set of likely outcomes.

These concepts were implemented in a human-interactive machine learning architecture (HIMLA) comprising four components (Figure 1): (1) a long short-term memory (LSTM) model that learns long-term cislunar orbital dynamics; (2) physics-informed training that incorporates cislunar dynamics through the loss function; (3) spectral and spatial learning fused with temporal features in a shared latent space; and (4) Bayesian inference that integrates human assessments with the fused machine-learning models to support co-learning. The approach is intended to improve detection and prediction accuracy while accelerating decision-making.

HIMLA was expected to provide more accurate cislunar sub-orbit location estimates than an LSTM alone. The LSTM was trained using an AFIT-provided cislunar dataset containing more than 400,000 data points, including velocity, position, time, and stability index (SI). The SI measures the difficulty of maintaining a given orbit and provides a useful indicator of potential mission intent. An SI of 1 represents a stable orbit suitable for surveillance; higher values indicate increasing instability, useful for more dynamic missions such as mobile surveillance disposal transfer.

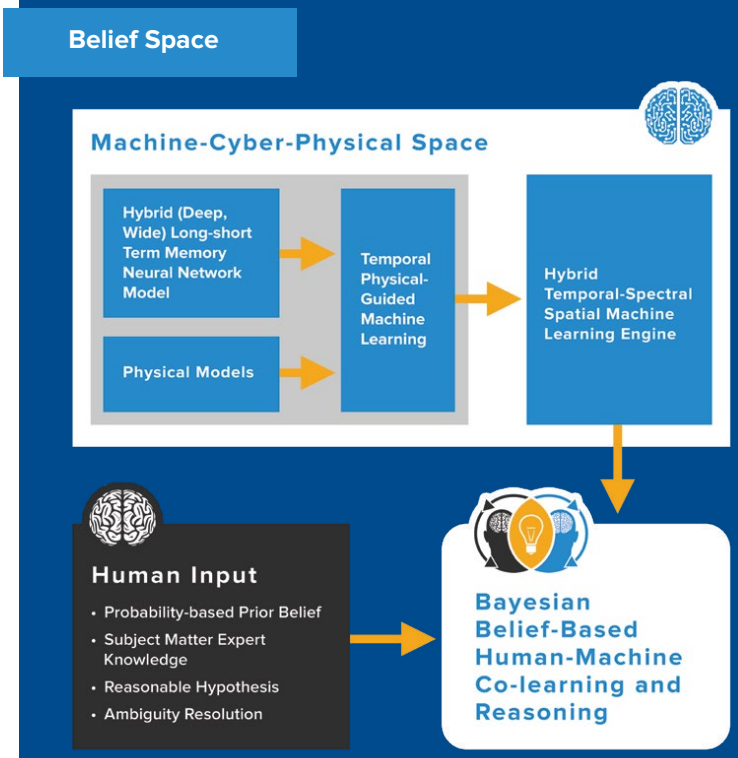


Figure 1. HIMLA workflows and belief spaces.

To evaluate HIMLA, LSTM predictions served as the Bayesian prior, human estimates of mission intent served as the likelihood, and the posterior represented the updated sub-orbit location prediction. Simulated event data and human estimates of spacecraft intent were used to test the model, demonstrating improvement over LSTM-only predictions (Figure 2). ☆

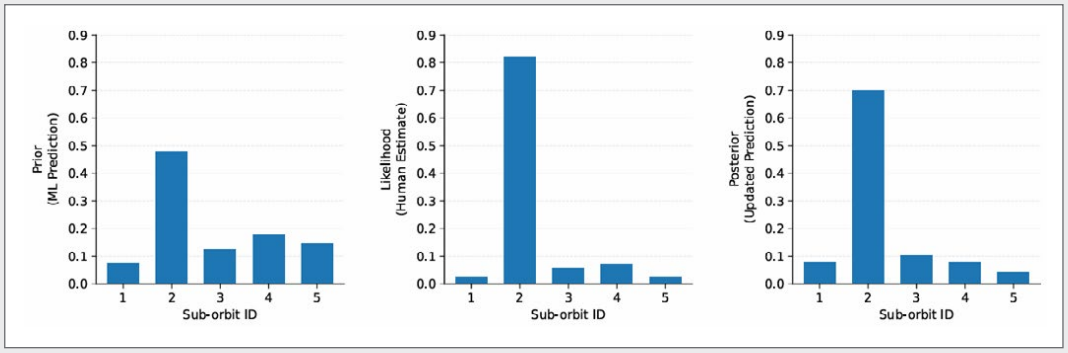
Dr. Lorraine Borghetti, Senior Research Scientist, 711HPW/RHWOH

1st Lt Michael Dillon, Computer Engineer, AFIT (formerly 711th HPW/RHWOH)

Capt Erjin Choi, Research Scientist, AFRL/RJSW

Graphics by 711 HPW/RHWOH

Figure 2. HIMLA Bayesian prior, likelihood, and posterior values for a representative cislunar scenario. The x-axis depicts the most likely sub-orbits (1–5). Right: LSTM-generated prior derived from recent sensor data. Center: likelihood based on simulated human operator assessments of spacecraft intent, informed by the SI. Left: posterior following Bayesian fusion of the prior and likelihood. Incorporating the human intent assessment increases the probability of sub-orbit 2 from 0.48 to 0.70, yielding a more actionable estimate for location-validation operations.



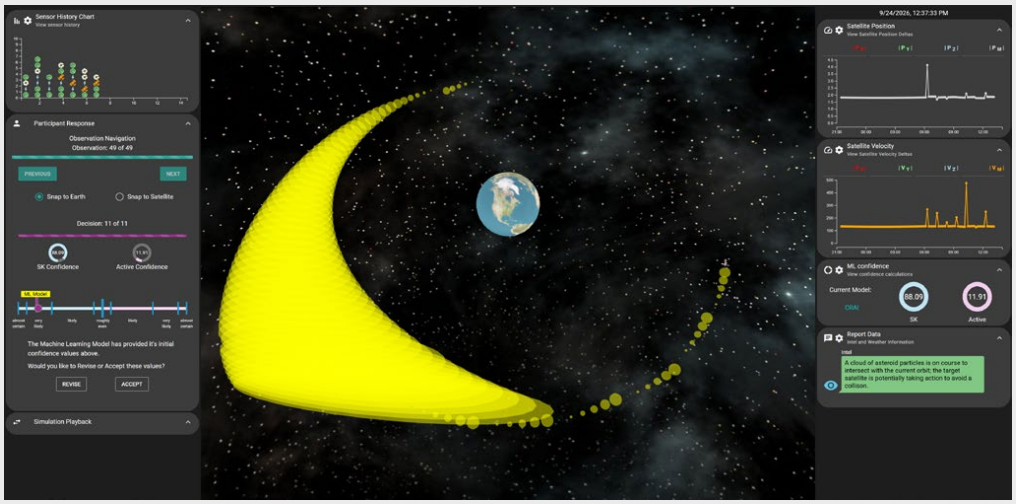


Figure 1. The Oko simulation environment to review simulated trajectory and contextual information to accept or revise the system's confidence estimates.

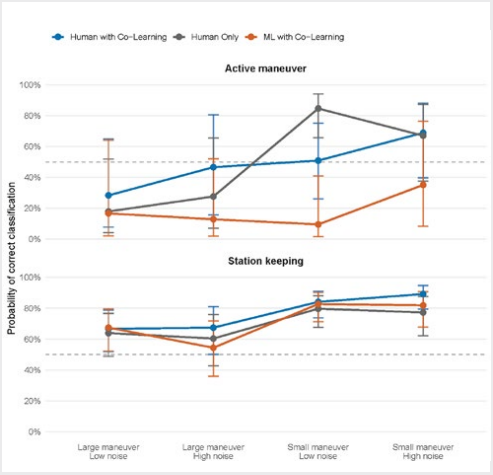


Figure 2. Comparisons of model-estimated correct-classification probabilities with/without machine-learning support, and the co-learning model output.

Screen capture by Mr. Allen Dukes, 711 HPW/RHWOH

Examining How Analysts and AI Learn Together in Space Domain Awareness

Assessing satellite behavior is a complex, uncertain problem. Human operators and artificial intelligence (AI) work together to accomplish this, but little is known about how their effectiveness can be maximized.

Previous work in our lab found that sensor noise, incomplete data, and accumulating observations could lead human-machine teams to incorrectly interpret normal variation in a satellite's trajectory as active maneuvering. In the present study, we replicated this experiment with a more advanced human-machine co-learning approach in a larger set of simulated space domain awareness scenarios.

Thirty-three novice participants reviewed simulated satellite trajectories and determined whether satellites were conducting active maneuvers or maintaining expected orbits, known as station keeping. Participants completed 20 scenarios involving varying maneuver sizes, sensor-noise levels, and relevant or non-relevant intelligence reports. One group had access to Co-learning Refined AI (CRAI) which is a machine-learning system that provided classifications and confidence estimates; a second group made assessments without the assistance of CRAI.

The study tested Co-learning by enabling users to accept or revise the CRAI recommendation. Those changes then informed later model predictions. The objective was not simply to determine whether AI could make classifications, but to examine how AI recommendations, human judgment, and feedback interact over time.

Results reinforced a key lesson from the previous work: machine-learning recommendations can strongly shape human decisions and human-machine teaming must be designed as a partnership rather than a handoff. Participants performed better when CRAI's recommendations were correct, but their decisions were less accurate when the recommendations were misleading. Although users could modify the system's suggested classification or confidence level, the model's underlying tendencies continued to influence their judgments. The study also found no consistent, dramatic overall performance advantage for the co-learning condition over the human-only group, underscoring that AI assistance alone does not guarantee better outcomes.

Furthermore, participants' confidence did not consistently align with their accuracy. In some cases, participants became more confident when they were correct, but also when they were incorrect. Machine-learning confidence estimates influenced users' reported confidence as well as their final decisions, demonstrating the importance of presenting AI recommendations in ways that support critical review rather than automatic acceptance.

One encouraging finding involved relevant intelligence reports. Contextual information significantly improved participants' ability to recognize active maneuvers, particularly when maneuvers were small or obscured by high sensor noise. The finding suggests that the most effective AI-enabled tools will pair computational recommendations with mission-relevant context for the human user.

The follow-on study also identified an important design challenge for co-learning systems. If an AI recommendation influences a user's decision, and that decision is then used as feedback to update the AI, errors could become self-reinforcing. Future research will examine ways to improve feedback mechanisms, help users detect faulty recommendations, and prevent inaccurate assumptions from becoming embedded in the human-machine team. ☆

Mr. Allen Dukes, Senior Software Engineer, 711 HPW/RHWOH

Dr. Lorraine Borghetti, Technical Lead and Co-Learning for Adaptive Human-Machine Teams LOE, 711 HPW/RHWOH

1st Lt Michael Dillon, Computer Engineer, 711 HPW/RHWOH

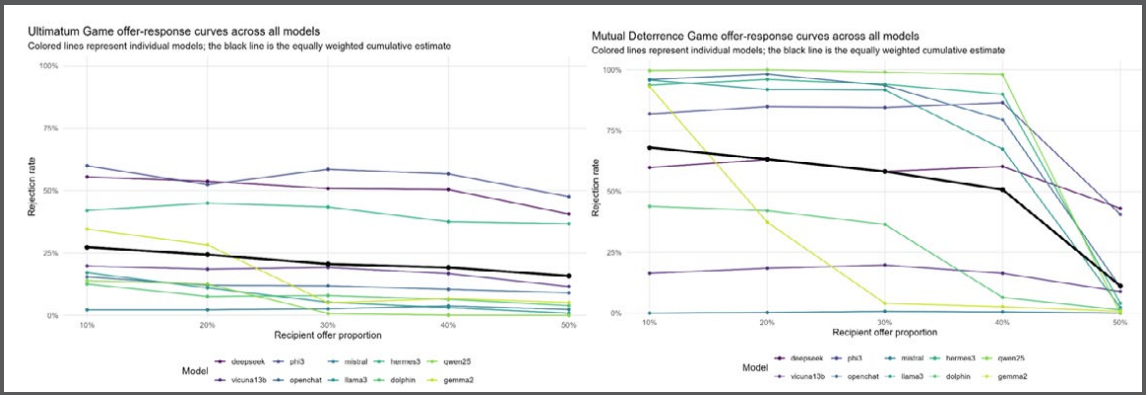


Figure 1. Retaliatory responses by LLM models depended on the fairness of the offer. In both UG and MDG, the models retaliated more frequently with more unfair offers. In UG, the model responses were insensitive across 5 out of 6 prompts, resulting in a weaker set of retaliatory decisions. Retaliation rates in the MDG reflected a more typical set of expectations, with higher retaliation with more unfair offers.

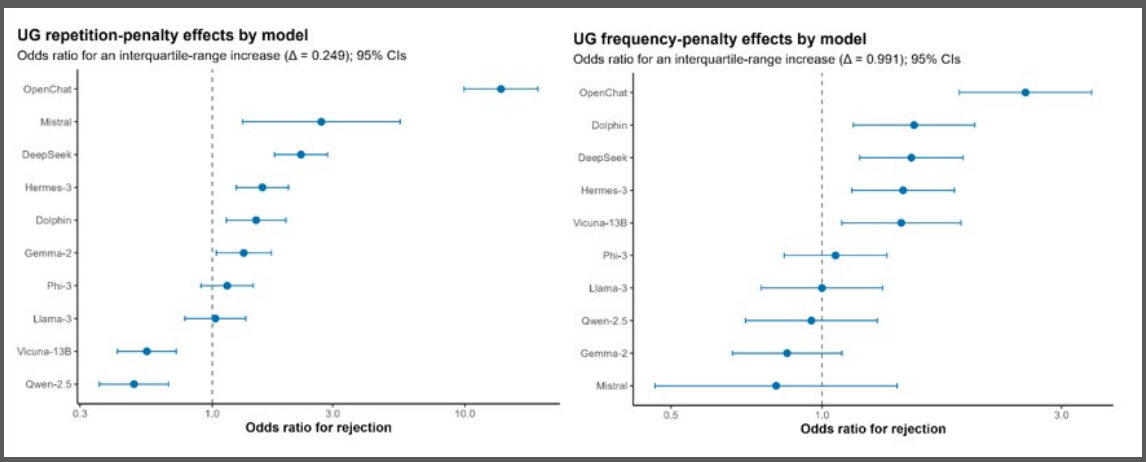


Figure 2. Retaliation rates varied depending on frequency, repetition-penalty, and model type. For some models, UG retaliation rates increased as repetition and frequency penalty increased, whereas for others retaliation rates decreased. These results suggest that benchmarking individual models in consequential decisions is important.

Charts by 711 HPW/RHWO

How LLM Settings Influence Retaliatory Decisions

Artificial intelligence systems built on large language models (LLMs) are rapidly moving from general-purpose chat applications into military, government, and commercial decision-support environments (e.g., Project Maven). In these settings, an LLM may not be asked to write a long response; instead, it may help a user or automated system make a binary choice, such as whether to retaliate, and such systems may be implemented with small models that exist on an isolated computer system. This effort was part of a co-learning approach to understand how intelligent systems using algorithmic approaches and game theory versus reasoning processes interact with human decision-making in deterrence scenarios.

Prior research indicates that LLMs can display escalatory tendencies. However, an important and less understood factor is the effect of the technical settings used to operate the model. LLMs are configured using many adjustable parameters that influence how predictable, varied, repetitive, or exploratory their outputs are. These settings are generally selected to improve user experience, but a configuration that works well for conversational tasks may not be suitable for a narrowly consequential decision.

This project examines how model configuration affects decisions involving costly retaliation. We tested ten distinct LLM architectures using two well-established decision scenarios [e.g., Ultimatum Game (UG); Mutual Deterrence Game (MDG)] in which a participant must decide whether to accept or reject an offer. Across more than 78,000 trials, we varied the amount at stake, the fairness of the proposed outcome, the wording of the instructions, and a range of model settings. The results show that LLM retaliatory behavior depends heavily on both the decision context and the way the model is configured. Across models, unfair offers generally produced more rejection and retaliation in both the Mutual Deterrence and Ultimatum Games (see Figure 1), consistent with established findings from human decision-making research.

However, the magnitude of the retaliatory response varied across models, prompts, and settings. Some adjustments such as repetition and frequency-penalty increased the likelihood that a model would reject an unfair offer, while other settings had stronger or weaker effects depending on the type of decision scenario and the model selected (see Figure 2). These findings suggest that LLM behavior should not be treated as fixed or assumed to be appropriate based solely on the model selected.

The Department of War considering LLM-enabled decision aids should test models against the specific choices they will be asked to support, identify behavioral biases or unintended escalatory patterns, and deliberately configure the system for its intended role. This work provides a repeatable method for benchmarking LLM decision behavior and for informing human review of how such systems are configured before they are relied upon in operational or other high-consequence contexts. ☆

Mr. Michael Chase, Research Scientist, 711HPW/RHWOH

Dr. Daniel Sazhin, National Research Council (NRC) Postdoctoral Fellow

Mr. Tyler Gandee, Associate Research Computer Scientist, 711HPW/RHWOH



Graphic by 711 HPW/RHWOH

MATTOCK: A Framework for Adaptive Open-Source Collection and Knowledge

Cognitive and multi-domain research for the Air Force and Space Force depends on large volumes of high-quality open-source intelligence (OSINT). Unrestricted web search, however, returns substantial noise and irrelevant content, making it difficult to assemble reliable corpora for modeling and analysis. Researchers need a collection approach that stays inside trusted sources while still adapting to varying and dynamic site structures across the open web.

Multi-domain Adaptive Text Traversal for Open-Source Collection & Knowledge (MATTOCK) addresses that requirement. It is an adaptive OSINT workflow that begins with curated seed sites and RSS feeds (i.e., trusted sources to initialize collection), retrieves fully rendered pages, and uses a Large Language Model (LLM) to generate site-specific extraction code. The result is a clean, reproducible corpus that retains page layout while reducing noise rather than a transient scrape. The same pipeline that produced the article dataset for the Co-Learning Line of Effort's Model of the Opponent (MOO) Project can be pointed at any domain simply by changing the framework's seed configuration.

Framework Components

Fetch and clean. A Playwright-driven node retrieves fully rendered pages, including JavaScript-heavy sites that static scrapers miss. Scripts, styles, and navigational noise are stripped before the content reaches an LLM, reducing token cost and focusing the model on article structure.

Adaptive extraction. Sending page HTML to an LLM for direct conversion to Markdown can clean up content, but it introduces the risk of generative hallucination and reduces precision. Instead, the LLM writes Python extraction code tailored to the page and domain. That code runs in a sandboxed subprocess and returns only content present in the HTML. The approach may miss some elements (resulting in lower recall), but it will not invent text that is not on the page (higher precision). Output quality is scored on word count, structural markers, and noise patterns. When extracted Markdown is considered low-quality (as defined by the user), code generation is retried with error context so the model can correct its extraction code.

Caching and persistence. Successful extraction scripts are cached by domain, collapsing latency and cost on subsequent visits. Extracted Markdown can be anonymized and stored in a corpus database, giving researchers a versioned, queryable store. The same graph supports both interactive single-URL use and continuous scrape/parse services.

Demonstration and Use Cases

MATTOCK's first production application was generation of the open-source article corpus that underpins ongoing MOO research, as well as supporting targeted data collection for the Weather Interference and Degradation (WIND) project. Because extraction is independent of any particular topic ontology, the framework is domain-agnostic. Air Force and Space Force teams can apply it to domains, e.g., space domain awareness, counterspace narratives, regional security, or other multi-domain lines of effort, by supplying an appropriate seed list. Near-term extensions under consideration include multilingual extraction and integration of pre-built prompts and models. Both aim to simplify applying MATTOCK to new research domains, while collecting from reliable sources defined by the user.

By constraining collection to known high-quality sources and adapting extraction to each site's structure without generative rewrite of the content, MATTOCK supplies the kind of clean, structured OSINT streams that cognitive and multi-domain modeling efforts require—without the noise floor of unrestricted search or the fidelity risk of free-form language-model summarization. ☆

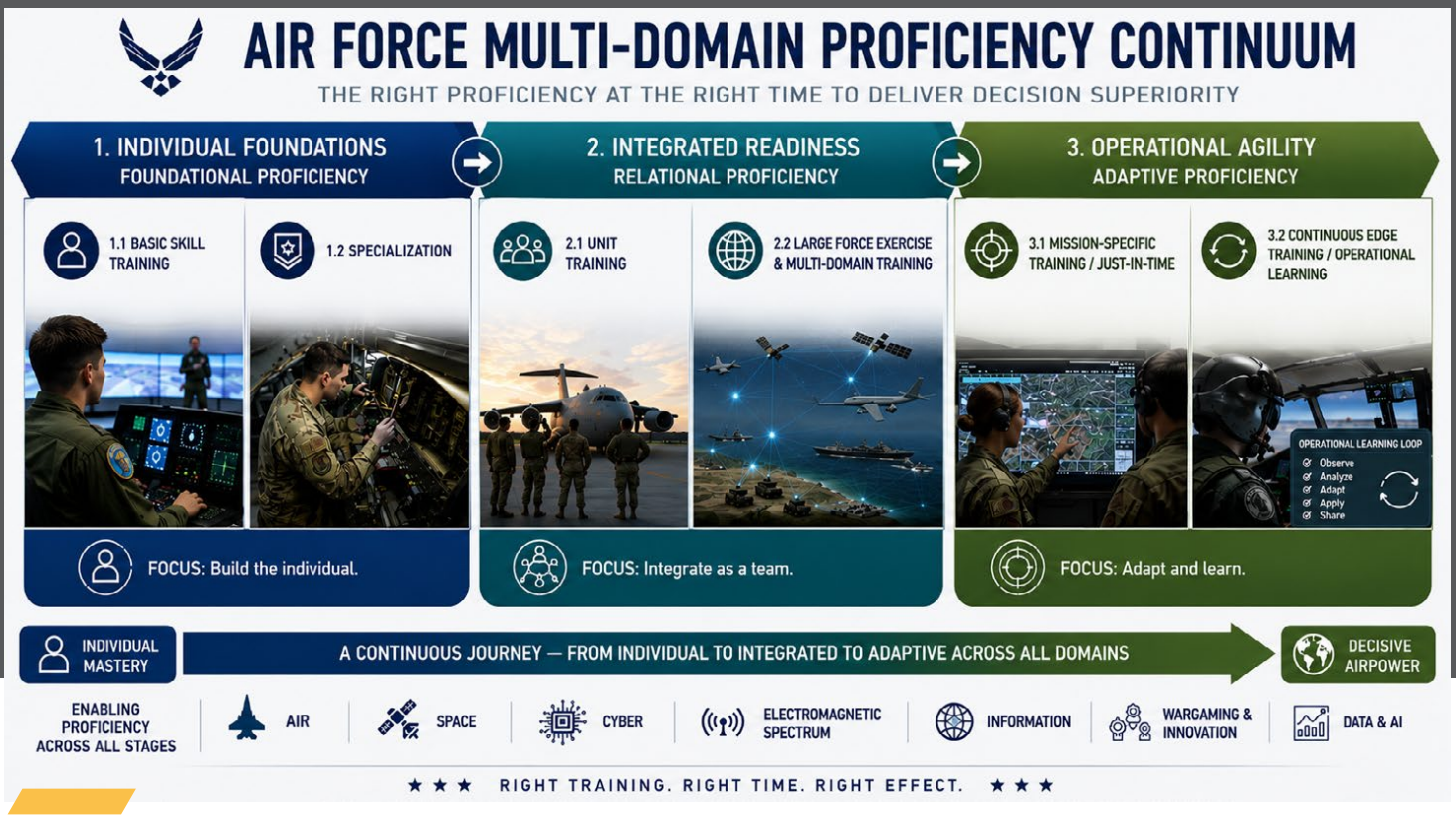
Mr. Tyler Gandee, Data Science Lead/Computer Scientist, 711 HPW/RHWOH

Dr. Jayde King, Research Psychologist, 711 HPW/RHWOH

TSgt Stephanie Vitkovitsky, WIND Co-Principal Investigator/Data Analyst, ACC 303 IS/DOC

SSgt Evan Hopper, WIND Co-Principal Investigator, ACC 303 IS/OMT

Figure 1. Air Force multi-domain proficiency continuum.



Graphic by Dr. Ryan Wohleber, 711 HPW/RHWOW

WARFIGHTER LEARNING TECHNOLOGIES

LINE OF EFFORT

The aim of the Warfighter Learning Technologies (WLT) line of effort (LOE) is to support capability throughout the Air Force multi-domain proficiency continuum (Figure 1) by researching, designing, and prototyping the warfighter learning technologies of the future. As we scale our LOE's capabilities to target the most challenging thrusts, we move forward focused on three major efforts:

1. Proficiency-based adaptive training
2. Adaptive and immersive artificial intelligence (AI) fluency training
3. Defining training requirements for collaborative combat aircraft (CCA)

Proficiency-Based Adaptive Training

Cognitive Assessment

This year, the WLT team has been developing cognitive assessment solutions to support Air Education and Training Command's (AETC) Comprehensive Readiness for Aircrew Flying Training (CRAFT). The team performed a deep dive into the Multi-Attribute Task Battery (MATB-II) assessment in various configurations and applications with partners at U.S. Army Aeromedical Research Laboratory (USAARL) and University of Dayton to thoroughly vet, orient, and augment it for formative evaluation in undergraduate pilot training. The team verified its applicability and identified opportunities to synthesize MATB-based assessment with another prominent approach – Model-Based Human-Systems Integration (MBHSI) – which breaks high-level operational tasks into a series of Basic Performance Resources (BPRs). WLT's efforts are making a key contribution to what promises to be a comprehensive and objective cognitive performance measurement system necessary to inform the application and configuration of targeted cognitive training that is a core mission of AETC's CRAFT program.

Program Leads

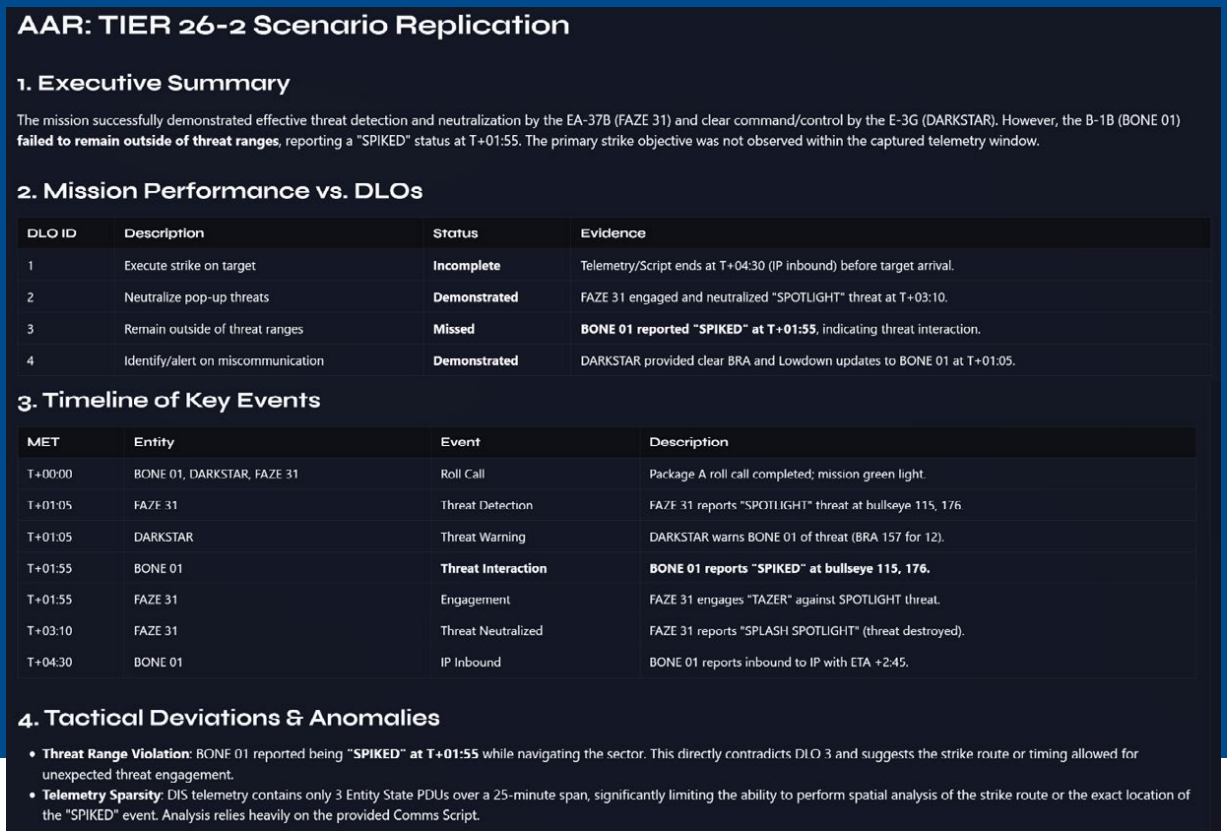
LOE LEAD
Dr. Ryan Wohleber

PROGRAM MANAGER
1st Lt Kyle Moats

Learner Modeling

WLT recently capped off its major partnership with the Air and Space Biosciences Division (RHB), the University of Cincinnati's Center for Sustainment of Trauma and Readiness Skills (C-STARS), the Army Combat Capabilities Development Command (DEVCOM), Vanderbilt University, and other collaborators to provide team-assessment of Combat Casualty Air Transport (CCAT) as both a product in its own right and as a proof of concept for a series of advanced analytical capabilities.

Continued on next page



AAR: TIER 26-2 Scenario Replication

1. Executive Summary

The mission successfully demonstrated effective threat detection and neutralization by the EA-37B (FAZE 31) and clear command/control by the E-3G (DARKSTAR). However, the B-1B (BONE 01) **failed to remain outside of threat ranges**, reporting a "SPIKED" status at T+01:55. The primary strike objective was not observed within the captured telemetry window.

2. Mission Performance vs. DLOs

DLO ID	Description	Status	Evidence
1	Execute strike on target	**Incomplete**	Telemetry/Script ends at T+04:30 (IP inbound) before target arrival.
2	Neutralize pop-up threats	**Demonstrated**	FAZE 31 engaged and neutralized "SPOTLIGHT" threat at T+03:10.
3	Remain outside of threat ranges	**Missed**	**BONE 01 reported "SPIKED" at T+01:55**, indicating threat interaction.
4	Identify/alert on miscommunication	**Demonstrated**	DARKSTAR provided clear BRA and Lowdown updates to BONE 01 at T+01:05.

3. Timeline of Key Events

MET	Entity	Event	Description
T+00:00	BONE 01, DARKSTAR, FAZE 31	Roll Call	Package A roll call completed; mission green light.
T+01:05	FAZE 31	Threat Detection	FAZE 31 reports "SPOTLIGHT" threat at bullseye 115, 176.
T+01:05	DARKSTAR	Threat Warning	DARKSTAR warns BONE 01 of threat (BRA 157 for 12).
T+01:55	BONE 01	**Threat Interaction**	**BONE 01 reports "SPIKED" at bullseye 115, 176.**
T+01:55	FAZE 31	Engagement	FAZE 31 engages "TAZER" against SPOTLIGHT threat.
T+03:10	FAZE 31	Threat Neutralized	FAZE 31 reports "SPLASH SPOTLIGHT" (threat destroyed).
T+04:30	BONE 01	IP Inbound	BONE 01 reports inbound to IP with ETA +2:45.

4. Tactical Deviations & Anomalies

- **Threat Range Violation**: BONE 01 reported being **"SPIKED" at T+01:55** while navigating the sector. This directly contradicts DLO 3 and suggests the strike route or timing allowed for unexpected threat engagement.
- **Telemetry Sparsity**: DIS telemetry contains only 3 Entity State PDUs over a 25-minute span, significantly limiting the ability to perform spatial analysis of the strike route or the exact location of the "SPIKED" event. Analysis relies heavily on the provided Comms Script.

Figure 2.
AAR review output from live exercise.

Screen capture by 711 HPW/RHWO

Learner Modeling (cont)

The completion of this program resulted in a sizable capability delivery including:

- Team Dynamic Measurement System (TDMS): software that uses dynamical systems metrics, AI, and communications algorithms to provide insight into the soft skills that underpin effective team function to empower and inform students and their instructors.
- Hierarchical Learner Modeling (HLM): an approach to competency modeling that affords capture of dynamic training data to systematically infer learner proficiency state.
- Video-based analytics: a computer vision approach that allows observation-based inference to inform HLM-based proficiency estimates via behavior and system interaction recognition.

Analytics for Large Force Exercises (LFEs)

This period, WLT supported the Testing/Training Information Evaluation and Research (TIER) LFE (iteration 26-2) as part of the Framework for Language Analysis and Resilience Enhancement (FLARE) capstone event by integrating and demonstrating a series of analytic capabilities for the exercise planning and execution team.

Automated After-Action Review (AAR) Summary and Query

Compiling an AAR is typically a painstaking and time-consuming exercise. WLT is addressing this through the development of an AI-powered assessment engine that provides well-organized overviews and actionable insights, alongside an interactive chat function allowing instructors to audit conclusions and query for additional information (Figure 2). This is achieved by deploying intelligent AI agents that use specialized tools to extract critical context from mission objectives and desired learning objectives, then synthesize that context with telemetry and communications data to produce a comprehensive appraisal of the training mission.

Network Analytics

Also demonstrated at the TIER 26-2 exercise as part of FLARE's capability was the application of network analytics visualizations to support sensemaking and allow instructors to rapidly identify communications phenomena and configurations that pose risk of, or result from, comms breakdowns in distributed teams. Demonstration of these analytics used hand-coded and verified utterances from LFE data with visualizations showing patterns of communication as assessed by these network analytics.

Modular Automated Behavior Envelope Analytics

WLT is developing prototype behavior-envelope analytics to help instructors and analysts make sense of complex exercise data and accelerate root-cause analysis for AAR preparation. By identifying an event's operational context (e.g., phase of flight, threat response, or communications challenge) the analytics assess observed actions against context-specific expectations. They provide time-stamped, graded insights that help instructors navigate significant events, identify potential risks, and inform assessments of individual and team proficiency. Results are displayed on an interactive timeline that shows contextual periods and associated graded events; users can select or create clips, configure supporting data feeds, and queue events for presentation (Figure 3 on next page).

Continued on next page

Screen capture by 711 HPW/RHWO

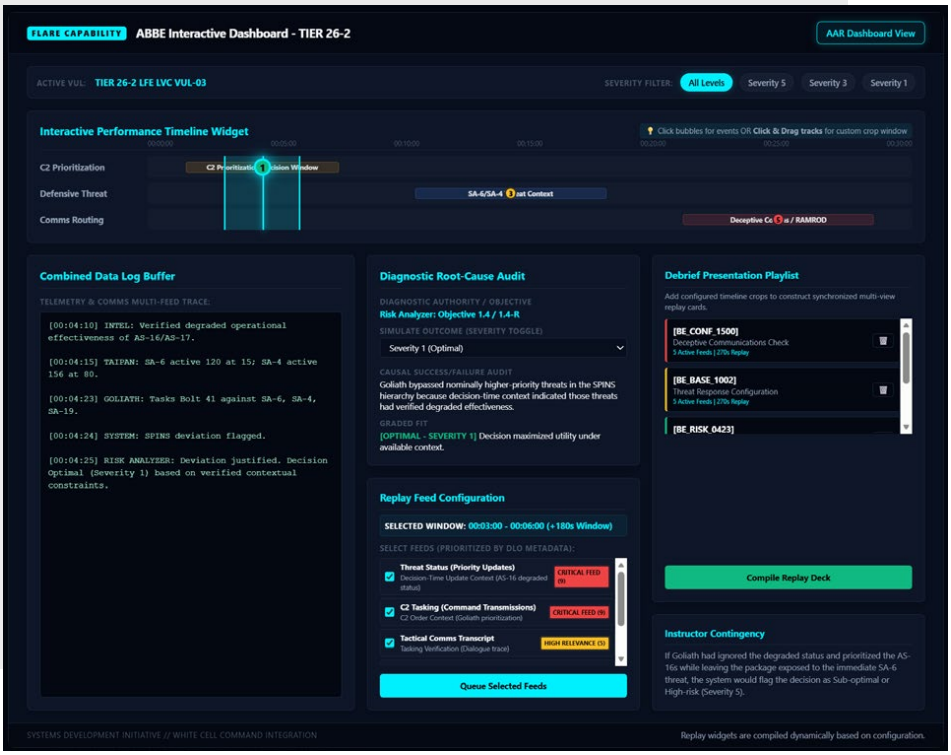


Figure 3. Behavior envelope-based analytics interactive AAR dashboard.

These analytics address the "it depends" problem when an action may be appropriate or inappropriate depending on the situation. For example, a trainee may prosecute a lower-priority threat when intelligence indicates that a higher-priority threat has been degraded. Fixed measures that compare behavior only to mission-planning expectations can misinterpret such decisions. Behavior envelopes establish the context needed to apply the relevant expectations, constraints, and tradeoffs.

Adaptive and Immersive AI Fluency Training

WLT is developing and evaluating an immersive, adaptive training approach to improve Warfighter's AI fluency, and specifically, their ability to use capabilities based on large language models (LLMs) as collaborative problem-solving partners, rather than simply as automation tools. AI literacy is an important starting point as it provides basic knowledge of prompting and known error patterns, but it does not by itself prepare users to assess AI outputs in open-ended, fast-moving work.

WLT's program initially focuses on discernment: recognizing potential inaccuracies, evaluating the plausibility of an AI system's reasoning, and investigating uncertainty when users cannot directly determine whether an output is correct. The approach to improving discernment is informed by the proposed tacit knowledge gap: the limitation that LLMs operate from written information while human expertise also relies on unwritten, experiential, and situational knowledge. WLT's framework uses realistic, ill-structured problems and the same AI and productivity tools users encounter in practice. By adapting scenario complexity to the learner's expertise and to the problem domain, the training is intended to develop three complementary capabilities: identifying output errors using domain expertise, assessing whether the AI's reasoning resembles expert reasoning, and exercising "active epistemic agency" which is the use of inquiry and verification to reduce uncertainty and guide the human-AI team toward a defensible conclusion.

Unlike conventional training that separates instruction, practice, and assessment, this training approach requires embedding structured calibration checkpoints within open-ended problem exploration. The design seeks to preserve the natural workflow of AI-supported problem solving while assessing learners' ability to calibrate trust, interrogate output validity, and manage uncertainty. WLT is developing domain-mapping methods, problem archetypes, and assessment checkpoints to test the framework's underlying claims, including whether experiential training improves transfer to new AI-interaction tasks relative to AI-literacy and other conventional instructional approaches.

Evaluating and Defining CCA Training

WLT is at the forefront of a critical research initiative to prepare pilots for the future of air combat operating CCAs through a project called Competency in CCA and Lethal Operations for Pilot Superiority (CICLOPS). This work focuses on understanding the training needs and cognitive demands fighter pilots will face when controlling unmanned platforms in addition to their own aircraft. Towards this goal, WLT has made several recent advances to understand and articulate requirements for training pilots to use the new capability including recent development of a survey distribution to pilots and Air Battle Managers (CCA Assignment and Operations Survey; CAOS) to gain insight about prospective CCA integration into pilot workflows.

In a partnership with the University of North Dakota (UND), WLT is developing unclassified components of the CCA research to identify underlying training and performance needs. Progress this year included the creation of pilot-vehicle interfaces (PVIs) and development of scenarios appropriate for civil aviation with a basis in operational requirements. This partnership provides naturalistic research opportunities to better understand the unique challenges the CCA paradigm presents, and to test and evaluate novel training approaches and applications of pedagogy to ensure readiness in these challenging environments.

Furthermore, the CICLOPS effort is actively aligning its simulation and training environments with broader, high-priority defense initiatives. By strategic benchmarking and mapping performance metrics from the wider fighter pilot and remotely piloted aircraft (RPA) communities, this work directly guides modifications to AFRL's native testbeds. This continuous strategic alignment ensures that current simulation infrastructure is precisely calibrated to support and validate next-generation warfighter technologies as they transition to the field. ☆

Dr. Ryan W. Wohleber, Warfighter Learning Technologies LOE Lead and Research Psychologist, 711 HPW/RHWOW









Program Leads

GRILL TEAM LEAD
Mr. Stephen McGee

PROGRAM MANAGER
1st Lt Aathi Anandan

For additional information, visit the GRILL's website:

WWW.AF-GRILL.COM

Photos by Mr. Will Graver (SierTek, Ltd.)

THE GAMING RESEARCH INTEGRATION FOR LEARNING LABORATORY®

GRILL® Vision

The Gaming Research Integration for Learning Laboratory (GRILL) leverages low-cost, high-fidelity, commercial off-the-shelf technology to create a variety of solutions for partners in the Department of Defense (DoD).

By pursuing these objectives, the team provides a greater return on investment for the DoD. Through utilizing existing technology to close training and simulation gaps, we reduce manpower costs by removing the need to develop tools from the ground up. Additionally, by utilizing platforms revolving around well-established game engines, the team ensures the continuity of our simulators by having tools that are frequently updated and compatible with emerging state-of-the-art hardware. The team's vision is to serve partners in industry, academia, and the DoD as a resource for the integration and application of game-based hardware and software to address Air and Space Force needs. By way of collaboration with our partners throughout the United States, the GRILL supports the development of training and simulation tools as well as a highly-qualified technical and scientific workforce.



Summer at the GRILL

During the summer of 2026 the GRILL hosted numerous Wright Scholars, teachers, United States Air Force Academy (USAFA) cadets, and interns. These participants worked collaboratively on community-driven challenge problems and developed free educational Science, Technology, Engineering, and Mathematics (STEM) content for teachers to use in the classroom.

The Wright Scholar program is enhanced by experienced engineers, real-world customers, and team-based learning, impacting their future STEM career opportunities and the broader community.

Summer Challenge Problems

1 BioEnvironmental Equipment Training (BEE)



This project aims to provide a realistic training module to airmen giving them detailed, guided instructions on how to use the RAE Multi Pro Gas Meter. The simulation offers several levels and scenarios in which airmen may potentially need to use the RAE Multi Pro Gas Meter to detect, gather readings on, and identify several types of hazardous and non-hazardous gases. The tutorial level gives detailed instructions to the user on how to calibrate and perform bump tests on the tool, before directing them out in a scenario to use the tool without any guidance. This project will be used to help train airmen to be multi-mission ready to use this tool without having the physical device beforehand due to shortages of the tool and guides, leading to more readiness and proper data tracking in hazardous situations.

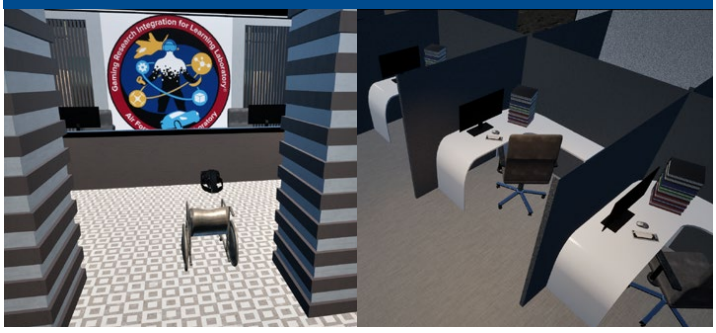
Screen captures by the GRILL / Photo by Mr. Will Graver (SierTek, Ltd.)

2 Crime Lab VR Tour



This project is a digital twin of the Montgomery County Coroner's Office and Miami Valley Regional Crime Lab that is accessible through desktop or virtual reality (VR). The goal of the project is to provide a tour through both facilities where users can do tasks that each lab has to perform themselves. In the digital twin, users will be able to choose which labs they want to explore and which activities in those labs they want to try. The program will teach them how to do each task step-by-step. This project will inform the general public about the crime lab and coroner's office and stimulate interest in forensics.

3 Accessibility Risk Immersion & Awareness (ARIA)



ARIA is based on creating a digital twin of a VR work environment (e.g., lab, office, workshop) from the ground up that is purposefully designed to allow able-bodied humans to notice environment changes that may hinder someone with a disability. The character is able to select from a wide range of disabilities. An office environment was designed with facilities and assigned daily tasks that allow the user (through VR with live physical feedback) to experience the limitations that a disabled employee may have to go through to highlight how this may affect the effectiveness and productivity of that employee - hindering performance.

Summer Challenge Problems

4 ISR Digital Experience (ISR-DX)



This project serves to simulate the strategy behind military intelligence, surveillance and reconnaissance (ISR) missions. The user has access to various ISR tools including imaging, signal, and cyber intelligence. The simulation requires the player to apply these tools to the game board in order to gather information and deduce the location of enemy points of interest (headquarters, cities, surface-to-air missiles, etc.). After identification, the player will form a strike path to the opposing main headquarters, attempting to avoid defenses. This project expands upon the physical version which relies on a time consuming pen-and-paper process and a human mediator to manually translate actions between players. Digitalizing the game eliminates the human middleman, drastically reduces gameplay length from its original 3-4 hours, and establishes focus on the game's core concepts. Eventually, the simulation will be applied to ISR classes in various universities and air bases, training analysts on the broader strategic scope of their work.

5 Ground Based Detection of SUAS (Small Unmanned Aircraft Systems)

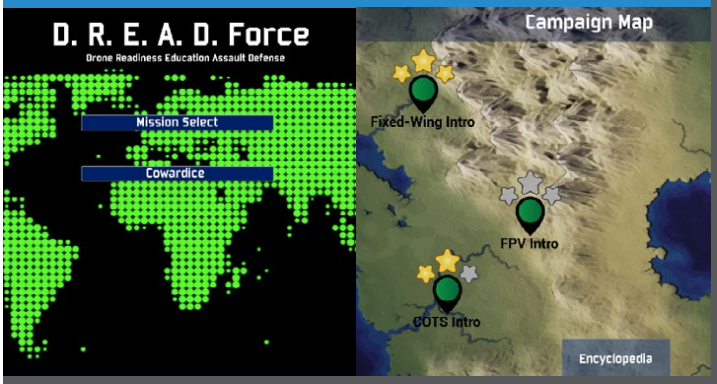


This project aims to deliver a realistic and immersive virtual simulation accessible through VR platforms. The system evaluates ground-based visual and acoustic drone detection scenarios under complex, real-world conditions like varied lighting, weather, and ambient noise. By providing precise control over drone altitude, speed, and sound, it effectively assesses participants' identification and tracking capabilities. It simultaneously logs key performance metrics, focusing primarily on the time disparity between auditory and visual detection. Ultimately, this simulation benchmarks human detection for behavioral research while providing a scalable training tool that eliminates the limitations of physical field testing.

Screen captures by the GRILL

United States Air Force Academy Cadet Summer Research Program

Drone Readiness Education Assault Defense Summary (DREAD)



The Drone Readiness Education Assault Defense (DREAD) counter-UAS learning game, developed for the United States Air Force Academy (USAFA), provides cadets with a new way to learn about drone threats and drone defenses. The game gives cadets missions to play through where they are given a Forward Operating Base (FOB) to defend. With limited resources, they have to protect their FOB against various incoming drone threats using the appropriate defense measures. The game will provide an interactive way for USAFA cadets to learn about various kinds of drone threats and proper ways to defend against each type of threat.

AI Physical Exam Summary



The AI Physical Exam, developed for the Uniform Services University of Health and Science (USUHS), offers students an innovative way to practice efficient physical exams. Through realistic scenarios using patients from incidents, the game challenges cadets to diagnose problems by selecting appropriate actions to test on patients. Students must analyze a patient's background, medical history, and personal attitudes to take specific actions to arrive at the final solution efficiently. Their performance is rated on both the quality and efficiency of their approach. This interactive platform provides USUHS students with valuable, interactive experience in managing diverse situations and medical incidents during physical exams.





Photos by the GRILL

Virtual Reality and Galvanic Vestibular Stimulation Research

The GRILL is running two research studies involving virtual reality (VR) and galvanic vestibular stimulation (GVS). The team is investigating how GVS can be used to reduce cybersickness and improve performance in VR environments by modulating vestibular input via non-invasive electrical stimulation. Although VR can be used as immersive and cost-effective training environments, cybersickness symptoms such as nausea, eye fatigue, and headaches can negatively impact performance and reduce training effectiveness. GVS may be a solution to mitigate cybersickness symptoms while positively impacting both user experience and performance.

The GRILL team first evaluated GVS in a VR Parachute Simulator. The operational trainer was developed in collaboration with the 336th Training Squadron to create an environment that could simulate a variety of malfunctions that could take place during a landing. The simulator replicated parachute landing in different scenarios (e.g., day and night, clear or overcast conditions) as the participants completed a series of jumps both with and without GVS simulating the angular descent as the participants attempted to land on a target. The researchers will assess the impact of GVS on cybersickness symptoms, workload, and landing performance metrics. Data collection ended in July 2026, with plans to share the research results soon.

The second study will utilize GVS in a flight simulator. Participants will perform a series of maneuvers ranging from basic turns and descents to more complex, aerobatic maneuvers such as a loop. The participants will execute the maneuvers with and without GVS, and the researchers will compare differences in flight performance, workload, and cybersickness symptoms. The flight scenarios were developed with fighter aircraft subject matter experts to confirm the maneuvers could be flown by participants with no flight training experience while also being engaging for experienced pilots. Data collection will occur during mid- to late-summer 2026. ☆



Human Factors XR Evaluations

The GRILL is supporting the Air Force by conducting a formal evaluation of emerging extended reality (XR) headsets. As future training environments demand more simulators within increasingly constrained physical footprints, many training sites are accelerating their transition to XR sleds. To guide this shift, the GRILL is performing a comprehensive, training task–based evaluation using the latest XR technologies across multiple image generators, sled configurations, and classification levels. This effort aims to identify current technology gaps and performance limitations in XR systems. The results will directly inform and shape the XR requirements for high end tactical flight training across multiple Air Force and joint service communities. ☆



Dr. Mark Draper
Warfighter Interfaces & Teaming
CTC Lead, 711 HPW/RHW

WARFIGHTER INTERFACES & TEAMING

CORE TECHNICAL COMPETENCY (CTC)

Operator interfaces are pervasive throughout the entirety of the Air and Space Forces; they provide critical communication connections between warfighters as well as between warfighters and machines, for all military missions. Teams of Airmen and Guardians employ warfighting power by optimally engaging and interfacing with increasingly complex and intelligent machines while under constant threats of situational complexity, data uncertainty, information subterfuge, and temporal constraints. Given that future wars will increasingly manifest as wars of cognition, victory will likely be decided by those who are best able to combine human cognitive abilities with increasingly capable machine capabilities to operate rapidly and effectively within the opponent's decision-making processes. Thus, optimized human-machine teaming (HMT) is required to win the next fight.

The Warfighter Interfaces and Teaming (WIT) CTC conducts research to enhance mission effectiveness across our Air and Space Forces by dynamically optimizing the integration of warfighter cognition with increasingly complex, autonomous and intelligent machines. WIT CTC research focuses on discovering, developing, evaluating, and transitioning adaptive warfighter interface technology, mission-optimized distributed team performance enhancements, agile communication management methods, and context-tailored intelligent decision aids/analytics to achieve and sustain decision advantage in complex environments. This CTC consists of two core research areas: 1) advanced human-machine interfaces [with emphasis on artificial intelligence (AI)-enabled systems], and 2) distributed teaming and communications.

Warfighter Interfaces & Teaming CTC

Distributed Teaming & Communication CRA

Human-Machine Interactions CRA

Automation and AI technology can be exceptionally powerful when operating within well-designed situations/environments, however they can rapidly become unreliable and unpredictable when operating outside of their competency envelopes. Current system design processes often solely focus on maximizing machine capabilities which unintentionally relegates the human role to serving as an ad hoc "catch all" problem solver: having to address any and all functions that the machine algorithms failed to accomplish. Thus, when unexpected contingencies arise, an out-of-the-loop warfighter is often expected to suddenly intervene and 'save the day' with little time and resources available to properly assess, understand, and act on the situation. This dilemma is often exacerbated by automation/AI designed with little-to-no explanatory ability as to its underlying functioning and reasoning (i.e., the 'black box' syndrome). A much-preferred approach to system design is to consider the human and machine as a joint-cognitive system from the beginning, where human and machine capabilities can be dynamically and smartly integrated to continually maximize performance across a range of realistic operating conditions and contingencies. This process also ensures that humans maintain meaningful human oversight and judgment of future AI-systems. WIT research aligns with this preferred approach to HMT design.

Continued on next page

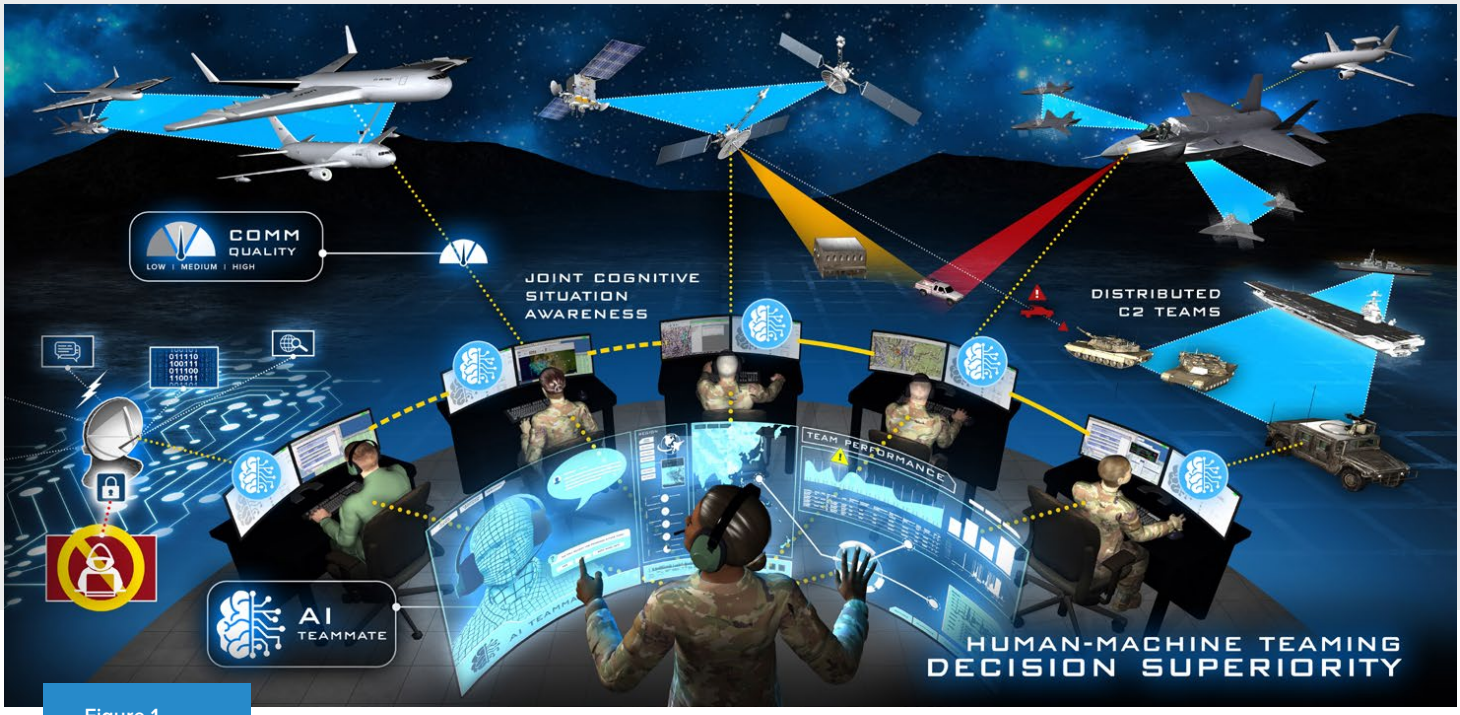


Figure 1

Graphic by Mr. Will Graver (SierTek, Ltd.)

In addition to the rapid expansion of machine capabilities, future fights will also be characterized by the need to 1) dynamically acquire and fuse information across multiple warfighting domains and 2) operate as increasingly distributed forces to effectively address adversary threats. Agile Combat Employment (ACE) and Joint All-Domain Command and Control (JADC2) concepts are premised on the ability to conduct agile multi-domain mission planning and execution activities across distributed heterogenous teams of warfighters and machines, and across C2 echelons (see Figure 1). Aggregating information across domains will create new information challenges (and opportunities) for the warfighter. In addition, communications between distributed forces will often be contested, degraded, or denied. Therefore, research is needed to better understand distributed teaming characteristics, challenges, and assessment methods to identify novel team coordination and collaborative solutions throughout a range of variable, degraded, and denied communication JADC2 environments.

CTC OVERVIEW

Conduct research to enable robust decision superiority across our Air and Space Forces by dynamically optimizing the integration of warfighter cognition with increasingly complex and intelligent machines/systems, creating maximally effective and resilient warfighting teams. Discover, develop, evaluate, and transition advanced adaptive warfighter interface technology, mission-optimized distributed team performance enhancements, communication management processes, and context-tailored intelligent decision aids/analytics to achieve and maintain decision superiority in uncertain environments against peer threats.

The WIT CTC is dedicated to conducting research that directly informs distributed teams and joint-cognitive systems by focusing on the highly complex, uncertain, and edge-case environments that will characterize future warfare. To further enhance its value proposition, the WIT CTC focuses on advancing and aggregating individual research projects into larger, more integrative solutions that enable high-priority Air & Space Force mission capability goals. Areas of specific concentration for the coming year include JADC2 battle management, layered base defense, the orchestration of non-kinetic effects with kinetic effects, and a variety of adaptive crewed-uncrewed operations.

So, what does the WIT CTC contribute to the fight? We provide contextually optimized interaction and teaming solutions that empower warfighters to team with each other and with AI to accomplish their missions more rapidly and effectively, thus providing robust decision superiority to our forces. Specific outputs of our research include advanced operator interface knowledge, guidelines and designs, teaming assessment metrics and collaboration aids, agentic AI teammates, agile communication enhancements, adaptive decision support, and advanced visualization tools. In the past few years, we have had multiple projects graduate to the next level of development including those related to C2 Battle management (JADPACT), team assessment (FLARE), and agentic agent decision aiding (BoBCAT).

The WIT CTC is executing specialized, warfighter-centric interface and teaming research that addresses high priority USAF needs and optimizes warfighter decision-making. With active projects focused on improving warfighter performance across Air, Space, Electronic Warfare (EW), Intelligence, Surveillance, and Reconnaissance (ISR), Cyber, and JADC2 environments, FY27 looks to be another exciting and productive year! ☆

Dr. Mark Draper
Warfighter Interfaces & Teaming CTC Lead and Principal Engineering Research Psychologist, 711 HPW/RHW



Dr. Michael Tolston
Distributed Teaming & Communication CRA Lead, 711 HPW/RHWT

DISTRIBUTED TEAMING & COMMUNICATION

CORE RESEARCH AREA



Future wars will be waged by distributed teams of warfighters and intelligent machines vying with adversaries for decision superiority in uncertain contested and denied environments. These distributed units will have to overcome significant teaming difficulties, including non-overlapping shared mental models, potential language and cultural barriers, limited knowledge of teammate skills and abilities, challenges associated with temporary team membership, as well as multilevel communications, coordination of disparate assets, and complex distributed decision-making. In addition to these fundamental teaming challenges, there will likely be real-time problems caused by denied or degraded communications that result in missing, incomplete, or incorrect information; delayed information; potential misinformation; information saturation and information bottlenecks; and fog of war.

To overcome these challenges, future teams will have to increasingly rely on technology that will enable faster team formation, resilient team performance, and more accurate situation assessment by ensuring common operational understanding across distributed settings to optimize real-time tactical adjustments for decision-making. In response to these demands, the Distributed Teaming and Communications Core Research Area (CRA) is focused on developing technology and methods to enhance the formation, assessment, and performance of distributed teams of warfighters, including human-human and human-autonomy teams. Our applications include ensuring decision superiority across our Air and Space Forces through research on distributed collaborative teaming and communications enhancements that facilitate resilient distributed teaming and rapid joint cognitive awareness and sensemaking.

Continued on next page

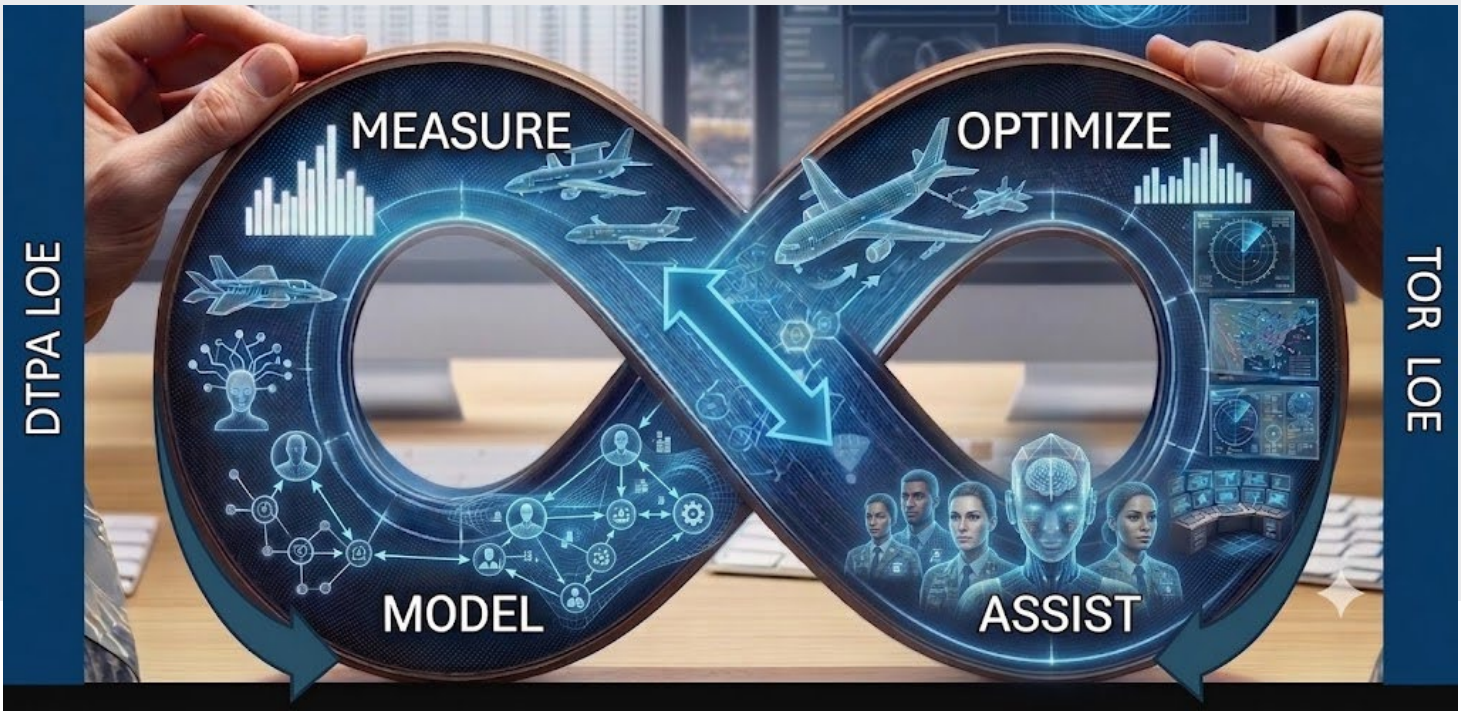
CRA OVERVIEW

Enhance the rapid formation, real-time assessment, and dynamically optimized performance of distributed, heterogeneous human-human and human-autonomy teams of warfighters. Enables resilient mission operations through the creation of novel distributed communication and collaboration tools, technologies, and AI-enabled teaming management.

Distributed Teaming & Communication
CRA

Dynamic Team Performance Assessment
LOE

Team Optimization and Recovery (TOR)
LOE



Graphic by 711 HPW/RHWT

The Distributed Teaming and Communications CRA is made up of two Lines of Effort (LOEs): Dynamic Team Performance Assessment and Team Optimization and Recovery.

Dynamic Team Performance Assessment (DPTA) LOE

The Dynamic Team Performance Assessment (DPTA) LOE is working to enable the rapid formation, real-time assessment, and dynamically optimized performance of distributed heterogeneous teams of warfighters and human-machine teams to promote rapid, agile, and robust mission operations. Research areas include factors that influence the formation of mission-effective heterogeneous teams (Teamwork Fundamentals, POC Ali Momen), dynamic monitoring of team performance via optimal assemblage of novel and existing metrics (Novel Team Performance Metrics, POC Gregory Funke), multimodal information processing for contextualized sensemaking (Contextualized Communication Machine Learning, POC Jeremy Gwinnup), and real-time contextual aids from team communication (Agile Communications, POC Michael Tolston).

Team Optimization and Recovery (TOR) LOE

To complement DTPAs measurement and modeling of teams, the Team Optimization and Recovery (TOR) LOE focuses on developing tools for distributed heterogeneous team coordination, collaboration, and agility to maximize team performance. The TOR LOE is focused on optimizing teams with intelligent aids, adaptive displays, and informed teaming interventions. Research areas in this LOE include designing, developing, and evaluating team optimization and recovery technologies to enhance communication, coordination, and decision-making among distributed teams (Team Collaboration and Resilience, POC Jessica Bartik), and conversational artificial intelligence (AI) technologies to enable high bandwidth natural communications with agent teammates (Tailored expressive Dialog Systems, POC Emily Conway; Synthetic Agent Voice Analysis Generation and Exploitation, POC Grant Erdman).

In sum, the work conducted in the Distributed Teaming and Communications CRA is designed to coherently integrate efforts across the sister LOEs to synergistically deliver technology that will enable warfighters to coordinate the integration of information from disparate sources, attend to a greater number of concurrent tasks through efficient teamwork, and collaborate across distributed teams in contested or denied environments. Expected outcomes from our research will empower enhanced teaming via dynamic measures and models of team performance and processes, dynamic interfaces and optimization strategies that make use of these models, naturalistic interactions with autonomous teammates, integrated cross domain interfaces, and robust AI-driven team assistance technologies. ☆

Dr. Michael Tolston, Distributed Teaming & Communication CRA Lead, 711 HPW/RHWT

DYNAMIC TEAM PERFORMANCE ASSESSMENT
LINE OF EFFORT

Program Leads

LOE LEAD
Dr. Gregory Funk

PROGRAM MANAGER
1st Lt Jacob Sampier

Dynamics of Team Insight:

Connecting Trust and Reliance to the Assessment and Improvement of Team Performance

A team's confidence in an artificial intelligence (AI) teammate changes as its actions unfold. Understanding how that confidence becomes willingness to rely, and eventually a shared decision, connects three papers and four experiments supporting the AFRL Dynamic Team Performance Assessment (DTPA) line of effort. The research identifies what to measure, how those measures change after failure, and what team dynamics reveal about subsequent reliance.

The first three experiments grew from AFRL's collaboration with the U.S. Air Force Academy (USAFA), combining behavioral-science expertise, cadet participation, and a shared platform for human–autonomy research. This partnership connects AFRL's Teamwork Fundamentals research to the practical challenge of assessing teams as they interact with increasingly capable AI.

Trustworthiness and Willingness to Rely

In the first experiment, published in Interaction Studies, 52 participants posed moral questions to an AI advisor through a text interface or a Furhat robot. Participants rated the advisor favorably on trustworthiness, averaging 6.28 on an eight-point scale, while intention to rely averaged 3.50 (Figure 1). Fifty-eight percent wanted explanations or justifications for its answers. For DTPA, the distinction establishes two complementary assessment targets: how people evaluate the AI and how ready they are to depend on it.

Following Recovery Across Missions

The second experiment examined what happens when that relationship is disrupted. Sixty-four USAFA cadets worked in two-person teams with an aerial AI teammate represented through Furhat. After the AI caused simulated civilian harm, a bundled explanation, justification, and repair message improved some moral appraisals without a significant gain in overall trustworthiness or intention to rely. Understanding the AI's account and regaining confidence in its future conduct emerged as distinct parts of recovery.

The third experiment followed 45 USAFA participants across five missions. After a common harmful decision in Mission 2, the AI either stopped or continued the violation. Corrected behavior produced more favorable recovery in trustworthiness from Mission 3 onward; intention to rely favored correction at Missions 4 and 5 (Figure 2 on next page). Timing matters for dynamic assessment: confidence and readiness to depend on a teammate can change at different points in the interaction.

Continued on next page

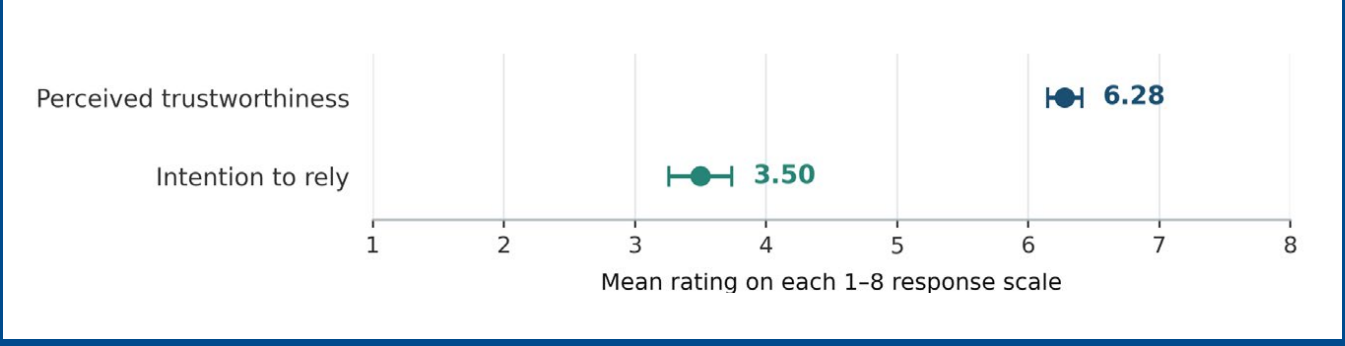


Figure 1. Favorable trustworthiness ratings coexisted with cautious willingness to rely. Dots show reported means; whiskers show ±1 standard error.

Source: Momen et al. (2025), Sections 4.4–4.5.

Chart by 711 HPW/RHWT

Trust Repair and the Pathway to Reliance

Graphs by 711 HPW/RHWT

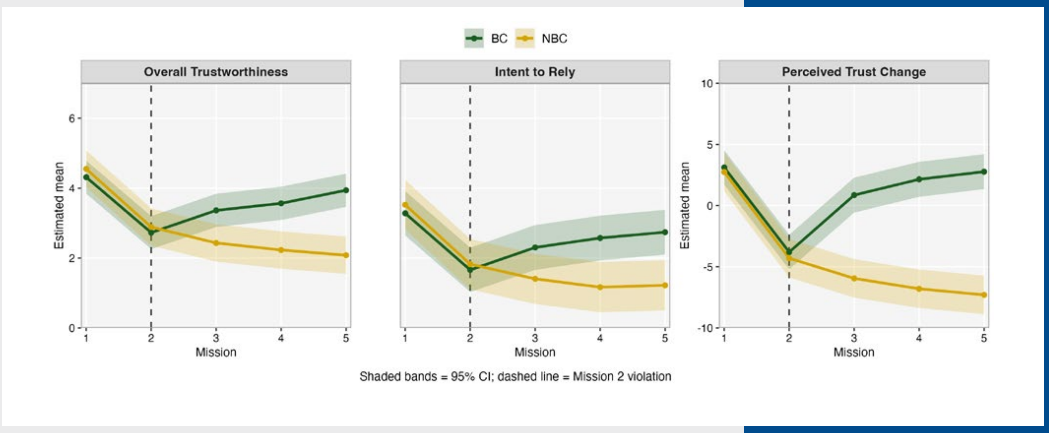


Fig. 2. Trust repair unfolds over time. Green: AI stopped the violation (BC). Gold: AI continued it (NBC). Bands show 95% confidence intervals; the dashed line marks the Mission 2 violation. Source: trust-repair manuscript, Figure 11.

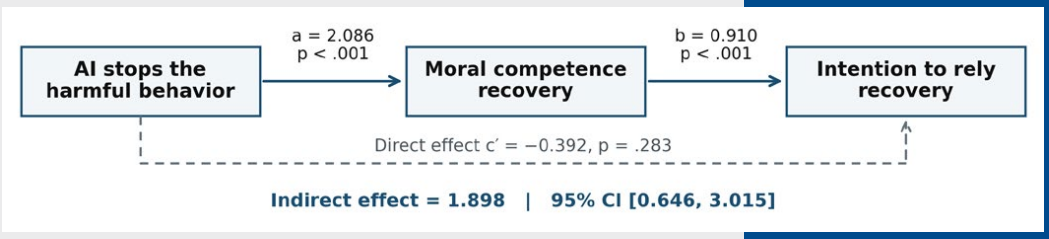


Fig. 3. Moral-competence recovery links correction to renewed reliance. Recovery is the Missions 3–5 mean minus Mission 2; *N* = 43. Solid paths are significant; the dashed direct path is not. Adapted from the trust-repair manuscript, Table 7 and Figure 13.

Mediation analyses connected this recovery to moral competence: participants' judgments of the AI's moral understanding and conduct. Moral-competence recovery statistically accounted for recovery in benevolence, integrity, and intention to rely, meeting the study's full-mediation criteria. Capability showed partial mediation. Figure 3 highlights the pathway to renewed reliance, giving DTPA a potential indicator of how confidence is being rebuilt.

Reading the Team as a Whole

The fourth experiment extends assessment to the team as a whole. Ninety-two university participants formed 37 teams of two or three people and completed a consensus prediction task with a NAO robot while researchers recorded cardiac activity. Teams then wagered shared points on the robot's performance in a new maze task.

An expanded analysis found that stronger whole-team cardiac coupling, measured through recurring joint patterns in teammates' heart-rate activity, was associated with higher-risk wagers. It added information beyond the agent-trust survey component. This connects a measure of the team's ongoing interaction to a subsequent collective reliance decision and advances DTPA's work on combining complementary team metrics.

Connecting Assessment to Better Decisions

Together, the experiments outline an assessment approach that follows the relationship between AI conduct, human judgments, and team decisions over time. Trust and reliance ratings describe individual responses; recovery trajectories reveal how those responses update; physiological dynamics add information about the team acting together. Combining these measures with task outcomes supports DTPA's goal of monitoring distributed human and human–machine teams and informing timely support.

We are now developing a fifth experiment in which AI actively helps the team make better decisions. The study will test whether that assistance improves collective decision quality. Building on the USAFA collaboration and the team-dynamics research, this next step connects insight into how teams develop and act on confidence with a direct assessment of the AI's contribution to team performance. ☆

Dr. Ali Momen, Research Psychologist, 711 HPW/RHWTE

Research Sources

Momen, A., Tossell, C. C., Niemeyer, R. E., Walliser, J., Tolston, M., Funke, G., & de Visser, E. J. (2025). Perceived trustworthiness and moral competence of a GenAI-enabled ethical robot advisor. *Interaction Studies*, 26(2), 326–356. https://doi.org/10.1075/is.25072.mom

Momen, A., Tossell, C. C., Lyons, J. B., Niemeyer, R. E., Cooley, K., Funke, G., Tolston, M., Schelble, B. G., Flathmann, C., McNeese, N. J., & de Visser, E. J. (2026). *Talk is cheap? Verbal repair falls short while behavioral change supports trust recovery in human–autonomy teams* [Unpublished manuscript]. Air Force Research Laboratory.

Tolston, M. T., Momen, A., & Funke, G. J. (2026). *Whole-team cardiac coupling predicts cross-task reliance on an agent teammate* [Unpublished manuscript]. Air Force Research Laboratory.

TEAM OPTIMIZATION AND RECOVERY
LINE OF EFFORT

Program Leads

LOE LEAD
Ms. Jessica Bartik

PROGRAM MANAGER
2d Lt Ryan Corner

Deployment of Synthetic Agent Voice Analysis Generation and Exploitation (SAVAGE) to the Rivet Joint Mission Trainer (RJMT)

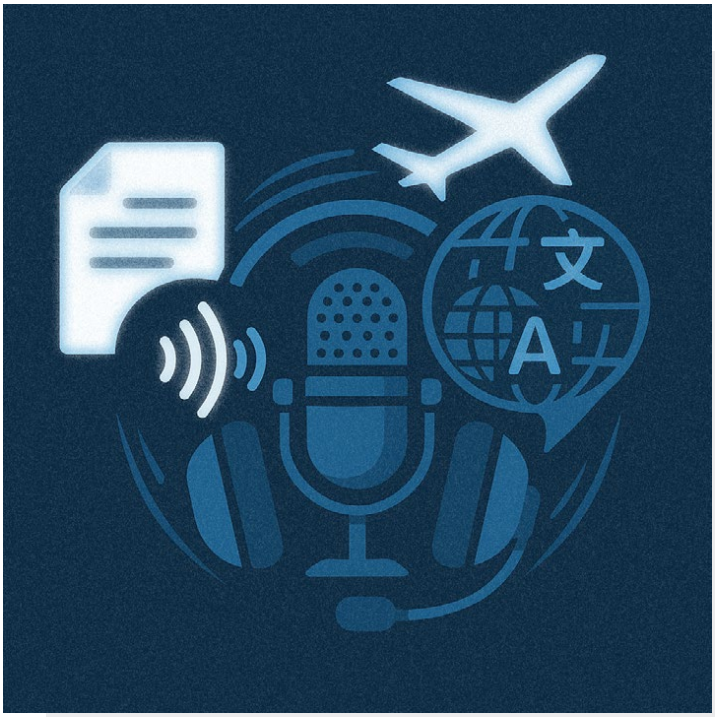
Speech Synthesis to Support Training of Cryptolinguists

A software container for speech synthesis was created and delivered to the 645th Aeronautical Systems Squadron (AESS) which supports the Rivet Joint Mission Trainer. SAVAGE allows multi-speaker, multi-lingual speech to be produced for training cryptolinguists. The speech synthesis tool allows a much larger and more variable speech content to be produced, supporting adaptive, high-fidelity training scenarios.

Hallmarks of the software are the inclusion of five languages and hundreds of speakers. The user can adjust the speech in its pronunciation, emphasis, and technical audio qualities. The resulting speech can be saved in over two dozen formats that are supported by the RJMT. This is accomplished through both command-line control and a graphical user interface (GUI).

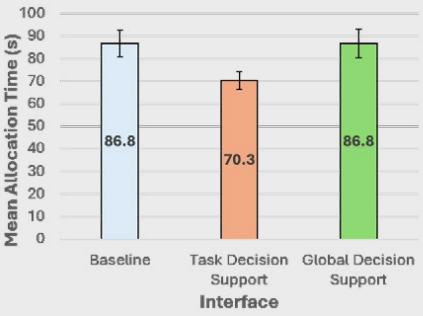
AFRL/RHW supplied a software container for consideration in the RJMT and assisted in adapting it to this trainer. The support contractors of the 645th AESS succeeded in including it in the RJMT, and it was deployed to users for beta testing in August 2026. Initial responses from the users indicate that the software has been well received and could be incorporated into the next block upgrade. ☆

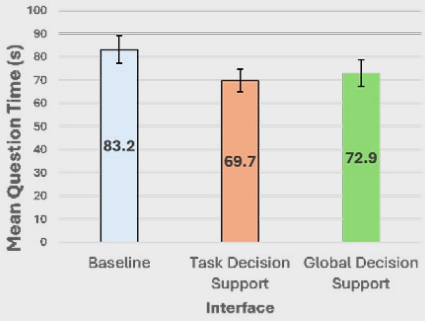
Dr. Grant Erdmann, Senior Research Mathematician, 711 HPW/RHWTE

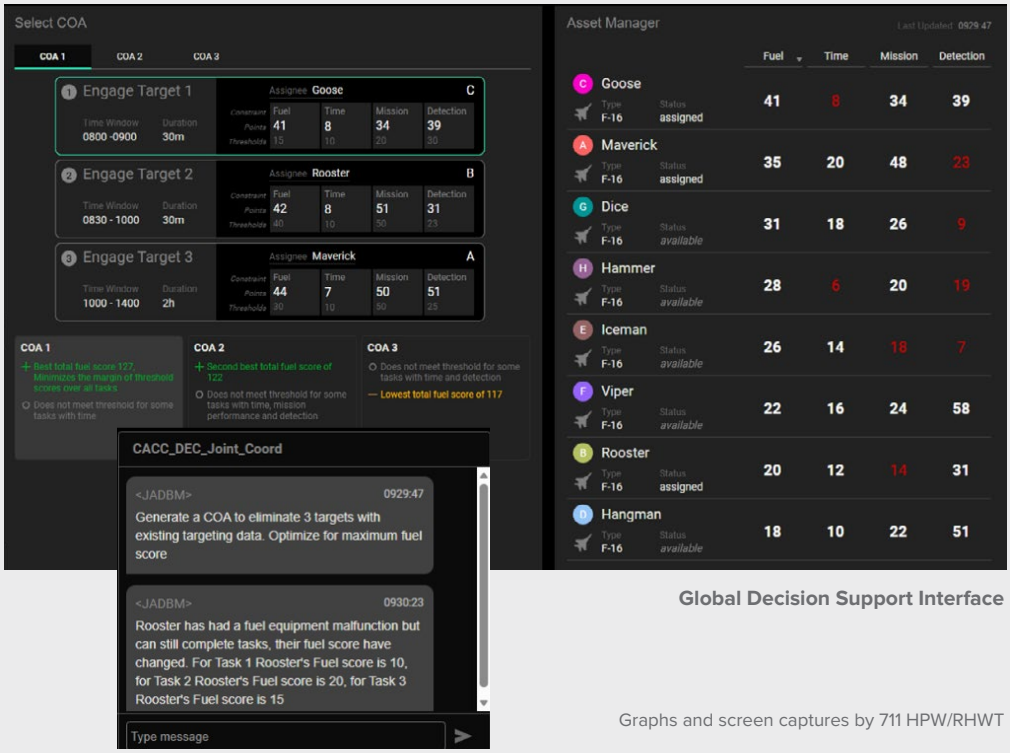




Graphic (above) by 711 HPW/RHWT
and photo by William R. Lewis







Global Decision Support Interface

Graphs and screen captures by 711 HPW/RHWT

Coordination User Interfaces for Joint Operations (CUJO) Study #1:

Decision Support Interfaces for JADC2 COA Generation

AFRL investigated the challenge of Air Battle Manager (ABM) asset allocation via the Team Collaboration and Resilience (TCR) project. To support Joint All-Domain Command and Control (JADC2) operations, ABMs must analyze a complex matrix of constraints, including relative asset location, speed, payload configurations, timing, fuel planning, risk, and probability of mission success. Currently, ABMs rely on Excel spreadsheets, manual calculations, mission requirements, and experience-based knowledge, creating a significant cognitive bottleneck. The TCR team modeled this complex ABM workflow as a hierarchical, multi-level optimization problem to inform the development of three interfaces to help ABMs generate asset allocation strategies to create more optimal courses of action (COAs). A study (N = 24) evaluated the utility of these three interfaces.

In this study, participants generated COAs that allocated one (of eight) F-16s to three tasks. Each task had thresholds for four mission constraints: Fuel Required, Time Required, Probability of Mission Success, and Probability of Detection. Participants were instructed to optimize for a specific constraint while also meeting the other three constraints. A *Baseline interface* presented data in a table format, with rows of eight aircraft (F-16s) and four columns of mission constraints. A *Task Decision Support interface* presented data with threshold color coding, a method to sort the aircraft based on a chosen constraint, and a parallel coordinates plot. A *Global Decision Support interface* presented the top three COAs generated by an artificial intelligence (AI) agent that considered aircraft data, mission constraint thresholds, optimization across the three tasks, and a COA comparison summary. Additionally, the *Global Decision Support interface* had the *Task Decision Support interface's* color coding and sorting features. The three interfaces were evaluated across two data states: when the presented data was true (Perfect) and when the data was stale (Imperfect). After the allocation was made, participants answered situation awareness (SA) questions.

Participants were faster with *Task Decision Support* than *Baseline and Global Decision Support* ($F(2,46) = 5.53$, $p = .007$, $\eta_p^2 = .19$). Participants were also faster ($F(1,23) = 55.66$, $p = .007$, $\eta_p^2 = .830$) and more accurate ($F(1,23) = 14.38$, $p = .001$, $\eta_p^2 = .39$) with Perfect Data than with Imperfect Data. Participants answered SA questions more slowly ($F(2,46) = 10.95$, $p < .001$, $\eta_p^2 = .32$) with *Baseline* than the other two interfaces. Participants were also slower ($F(1,23) = 20.29$, $p < .001$, $\eta_p^2 = .47$) and less accurate ($F(1,23) = 9.63$, $p = .005$, $\eta_p^2 = .30$) with Imperfect Data compared to Perfect Data.

Results indicated that the *Task Decision Support interface* improved participant allocation time without sacrificing accuracy (rank). Participant comments indicated that they did not trust the *Global Decision Support interface's* AI agent's recommendations. This may be due to limited time with the tool and lack of transparency into the AI's reasoning. These results provide actionable guidance for the design of future decision-support interfaces to aid ABMs in supporting JADC2 operations. ☆

Ms. Jessica Bartik, Research Psychologist, 711 HPW/RHWT
Dr. Hunter Oldham, Research General Engineer, 711 HPW/RHWTC
Mr. Dakota Evans, Mechanical Engineer, 711 HPW/RHWTC
Mr. Heath Ruff, DCS Corp
Dr. Kyle Behymer, DCS Corp
Mr. Andrew Eckert, DCS Corp

DYNAMIC TEAM PERFORMANCE ASSESSMENT AND TEAM OPTIMIZATION AND RECOVERY

CROSS-LOE COLLABORATION

Contextualized Communication Machine Learning (CCML) in Partnership With Synthetic Agent Voice Analysis Generation and Exploitation (SAVAGE)

Speech Recognition Accuracy Study in Support of CCA Program of Record

A study to examine the accuracy of Automated Speech Recognition (ASR) systems using a high-environmental fidelity simulator Informed Collaborative Combat Aircraft (CCA) Program of Record (PoR) of accuracy of State-of-the-Art speech recognition technology for their specific domain. This informed the CCA PoR of the accuracy improvements possible with additional tuning of the models. The CCA PoR will use this information on future designs and work to expand requirements on speech-based approaches to CCA control.

This work was a collaboration with RHB and collected representative CCA control speech in the 711 HPW Research Altitude Chamber (RAC), a cockpit-based altitude chamber, with subjects wearing full kit (mask) at altitude and with noise conditions in coordination with CCA PoR.

Six Commercial ASR systems were evaluated on the representative data collected from the RAC where word error rate (WER) ranged from 28% to 9.7%. Then two different commercial ASR models were tuned with AFRL AI Compute resources using synthetic speech modeling the CCA commands by utilizing the Text-To-Speech technology coming from SAVAGE. These tuned models reduced the WER to 1.5% on average showing that there are existing methods to accurately recognize CCA command specific speech in the cockpit.

AFLCMC/WA (CCA PoR) plans to utilize these results to inform decisions on future interface development on CCA control. Also, the tuned ASR models developed under this effort have been transitioned to the DARPA VENOM program office for their use. ☆

Dr. Jeremy Gwinnup, Senior Research Computer Scientist, 711 HPW/RHWTE

Dr. Grant Erdmann, Senior Research Mathematician, 711 HPW/RHWTE

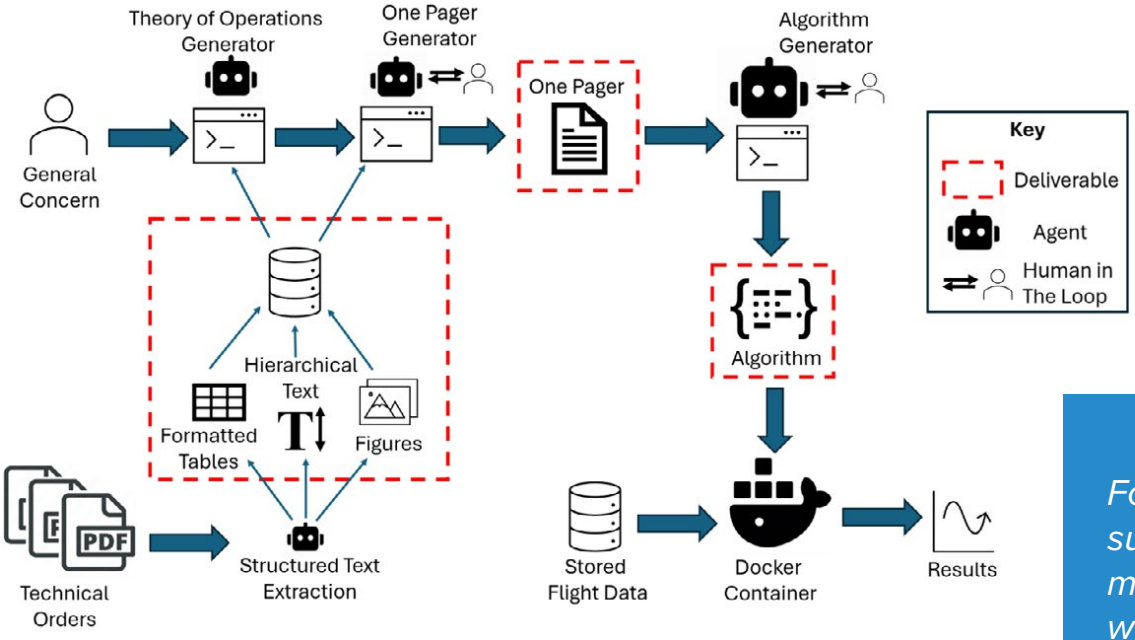




Photos (top) by Megan Gallo, 711 HPW and (bottom) Rick Eldridge, 711 HPW CAG

BEST OF BREED CONFIGURABLE ADAPTIVE TEAMMATE (BOBCAT)

CANDIDATE PRODUCT



Graphic by 711 HPW/RHWT

Forging decision superiority across military domains with advanced human-machine teaming

BoBCAT in Support of Condition Based Maintenance Plus (CBM+)

The Best of Breed Configurable Adaptive Teammate (BoBCAT) is a 711th Human Performance Wing (AFRL/711 HPW/RHW) research program developing Agentic Artificial Intelligence (AI) as a flexible and configurable teammate for military-relevant domains, delivering evaluated software prototypes/products to enhance warfighter decision-making through intuitive human-machine teaming (HMT). BoBCAT directly responds to the DAF AI Strategy Mission Area 1, AI for Decision Superiority in Multi-Domain Operations, addressing a gap associated with the ingestion/interaction with large, varied datasets and the necessary assistance to conduct meaning-making on that data. BoBCAT will act as an enabler for systems that rely on manual processes to have a framework for integrating AI/machine learning (ML)/large language model (LLM) based agents, acting as a force multiplier for analysis, planning, and execution.

An Agentic AI workflow was built to assist the Air Force Life Cycle Management Center Rapid Sustainment Office (AFLCMC/ROD) in their mission supporting Condition Based Maintenance Plus (CBM+). Their mission has been "BoBCAT enabled" so their operators can more efficiently and effectively conduct their mission. This work has mainly focused on the "BoBC" part of BoBCAT.

Step 1 calls an Agent to create a structured data index from the relevant Technical Orders (TOs) containing navigable information of Sections, Tables and Figures.

Step 2 allows the User to input a General Concern, and an Agent suggests multiple Theory of Operations based on information in the TOs.

Step 3 uses multiple Agents which take the Theory of Operation, User input, and the extracted TO content to generate a "One Pager" as defined by the RSO. This is an iterative human in the loop process with subject matter expert (SME) feedback. This "One Pager" is sent to a large Code Generation Agent that produces an algorithm to search through the Flight Data information. This generated algorithm can be run against the large set of Flight Data Recorder (FDR) data and will generate plots for all detected anomalies.

Overall, this is a great example of a Team of Agents working through a practical, mission specific problem and we are generating results in minutes instead of days/weeks. ☆

Mr. Eric Hansen, Senior Electronics Engineer, 711 HPW/RHWTE



Mr. Peter Venero
Human-Machine Interactions
CRA Lead, 711 HPW/RHWI

HUMAN-MACHINE INTERACTIONS

CORE RESEARCH AREA



The Human-Machine Interactions (HMI) Core Research Area focuses on how operators interact with systems, ranging from Pilot Vehicle Interfaces (PVI) to advanced automation and artificial intelligence (AI)-enabled decision support and analysis tools. In short, our research investigates how operators interact with a single system or component of a larger system, rather than large-scale human-machine teams. Indeed, ours is a vital endeavor to achieve large scale Joint All-Domain Operations (JADO). Our CRA members are experts in human factors design, sensory perception, knowledge elicitation, decision-making, data representation, and trust and transparency. Application areas include next-generation cockpit interfaces, exploring adaptive systems that adjust their behavior and modify interfaces based on operator states, developing human-centric tools for intelligence analysis, applying design guidelines to establish trustworthy and transparent intelligent systems, and solving the problem of transforming 'big data' into intuitive, quickly comprehensible representations that can be leveraged by operators. Lastly, the CRA explores novel interface concepts for Cognitive Warfare applications. The 711 Human Performance Wing (711HPW) has prioritized Cognitive Warfare as a thrust area, and a portion of our portfolio reflects this new investment area. ☆

Mr. Peter Venero, Human-Machine Interactions CRA Lead and Senior Research Psychologist,, 711 HPW/RHWI

CRA OVERVIEW

Identify principles of human interaction with highly complex systems, including advanced automation and increasingly intelligent AI-enabled machines. Achieve and sustain decision superiority and performance across complex and uncertain mission environments. Identify, characterize, and overcome key challenges to warfighter interactions with complex and intelligent systems such as situationally-adaptive interface design and usability, knowledge representation across sensory modalities, system observability and transparency, directability, joint cognitive decision-making, and maintaining calibrated trust across changing conditions.

Human-Machine Interactions
CRA

HMI-enabled Decision Superiority
LOE

Rapid Joint Cognitive Awareness
LOE



Graphic by Mr. Will Graver (SierTek, Ltd)

HMI-ENABLED DECISION SUPERIORITY

LINE OF EFFORT

Operational Impact of Non-Kinetic Effects (OINK)

With emerging advancements in artificial intelligence (AI) and autonomous systems, the modern battlefield landscape is rapidly becoming more complex and interconnected. As technologies continue to evolve, current and future military operations will require deliberate synchronization of kinetic and non-kinetic effects (NKE) to ensure a joint common operating picture (COP) and achieve mission success.

Kinetic operations involves the application of direct, physical force to destroy an adversaries infrastructure and operational capabilities. In contrast, non-kinetic operations leverages space, cyber, and electromagnetic domains to deny, degrade, disrupt, or destroy (D4) existing infrastructure to shape the operational environment without relying on physical destruction. Yet, the ability of command and control (C2) nodes to effectively synchronize and execute the integration of kinetic and non-kinetic effects at scale remains an ongoing challenge.

The Human-Machine Interface Enabled Decision Superiority (HEDS) line of effort is tackling this constraint head on by identifying current bottlenecks and pain points collected from cognitive task analysis (CTA) directly from the warfighters while simultaneously developing three non-kinetic capabilities postured at enhancing situational awareness and battlefield course of action (COA) for the non-kinetic duty officer (NKDO) at air operation centers (AOCs):

Program Leads

LOE LEAD
Mr. Peter Venero

PROGRAM MANAGER
Capt Brody Jones

Consolidated Operational Perspective of Cyber Operations (COPCO) is a capability focused on displaying critical information collected and obtained through cyber network operations and is aligned with tactical operations from the Combat Mission Teams (CMTs). This effort is led by Mr. Peter Venero. Space Power Exploitation for Actionable Response (SPEAR) is a capability to track and monitor space domain awareness (SDA) while enhancing non-kinetic situational awareness, decision-making, and mission planning. SPEAR is aligned to accelerate space duty officers (SDO) workflow process. This effort is led by Dr. Justin Nelson. Scalable Tactical Overlays for Electronic Warfare in Air and Space (STOEWAS) is a capability to address persistent challenges in the electromagnetic spectrum. STOEWAS is a mobile technology that will be integrated within airborne vehicles including Compass Calls, Rivet Joints, Growlers, and Raiders. This effort is led by Dr. Chris Brill.

Individually, each capability operates as a standalone technology to enhance situational awareness and mission efficacy within its respective domain. Collectively, intelligence within each capability will be aggregated and disseminated into a single user-interface 'Gravity Well' to provide a complete COP to the NKDO to provide us an edge in the battlefield. ☆

Mr. Peter Venero, NK Cyber Lead and Senior Research Psychologist, 711 HPW/RHWI
Dr. Justin Nelson, NK Space Lead and Senior General Engineer, 711 HPW/RHWIM
Dr. Chris Brill, NKE Warfare Lead and Senior Research Psychologist, 711 HPW/RHWIM

Photo by ACC 55th Wing



Members of the RHWI STOEWAS team with leaders of the ACC 55th Wing Det. 1 at Davis-Monthan Air Force Base.

Scalable Tactical Overlays for Electronic Warfare in Air and Space (STOEWAS)

STOEWAS (pronounced stow-was) is a new technical effort under the HEDS LOE, serving as one of three non-kinetic pillars alongside space and cyber. The program focuses on developing operator-centric Human-Machine Interfaces (HMIs) for Electronic Warfare Officers (EWOs) to address the persistent challenge of maintaining battlespace awareness within the electromagnetic spectrum.

Designed as a tactical planning and operations aid, STOEWAS will utilize a portable form factor (such as a tablet). This mobile approach offers operational flexibility while bypassing the costly, complex process of integrating new software "under glass" on existing cockpit displays.

Target platforms for STOEWAS integration include:

- EA-37B "Compass Call"
- RC-135V/W Rivet Joint
- EA-18G "Growler"
- B-21 Raider (exploratory focus, leveraging its electronic attack capabilities to operate in anti-access/area denial zones)

Currently in its first year, the project is establishing a foundation rooted in human factors design. This research builds upon a prior tradecraft analysis executed by the team for Ephemeral Paragon (E-Gon)—an AFRL advanced development program focused on next-generation adaptive software for electronic attack and electronic protection.

STOEWAS extends this foundation by executing cognitive task analyses with operators. These analyses capture critical decision requirements, workflows, communication patterns, existing tools, pain points, and unmet cognitive needs. The team uses these empirical insights to generate novel interfaces, gather operator feedback, and rapidly iterate designs.

To initiate operator touchpoints, the team attended the Spectrum Integration Group (SIG) at Nellis AFB, strengthening connections with the Electromagnetic Spectrum Operations (EMSO) community.

The team has since executed targeted field activities across key operational units, starting with the 55th Wing Detachment 1 (Davis Monthan Air Force Base; Tucson, AZ), where they observed a planning exercise, a simulated mission, and conducted cognitive interviews with EWOs. Site visits to additional operational units are projected for FY27.

Although STOEWAS is designed as a tactical Electronic Warfare (EW) tool, key elements will feed directly into Gravity Well—an operational software tool designed to help Non-Kinetic Duty Officers (NKDOs) seamlessly integrate non-kinetic facilitators into kinetic-centric operations. To better understand the interplay between tactical non-kinetic operators (e.g., airborne EWOs, aka "CROW") and NKDOs, the team executed data collections with the 612th Air Operations Center (AOC; Davis-Monthan AFB, AZ) and the 603rd AOC (Ramstein AB, Germany). While at the 603rd, the team observed an international live-fly exercise from AOC/A39 – information operations – offering a keen perspective of NKDOs and how they view tactical EW assets.

Lastly, the STOEWAS team coordinates closely with Headquarters Air Force (HAF) A39. HAF/A39 is exploring an Electromagnetic Battle Management (EMBM) concept with Air Combat Command and has leveraged data from both the STOEWAS effort and the legacy E-Gon study. This continuous consultation ensures robust information sharing and deconfliction. In return, Headquarters Air Force (HAF)/A39 provides the STOEWAS team with vital access and connections to the broader operational community. In FY 2027, the team will conclude its field data collection and pivot fully to developing and iterating prototype software interface designs. ☆

Dr. J Christopher Brill, Senior Research Psychologist, 711 HPW/RHWIM



Graphic by Mr. Will Graver (SierTek, Ltd)

RAPID JOINT COGNITIVE AWARENESS (RJCA)

LINE OF EFFORT

Artificial intelligence (AI) enabled systems have become critical tools for Intelligence, Surveillance, and Reconnaissance (ISR) analysts. However, there are no validated frameworks nor established strategies to share tasks between human operators and teamed AI systems. The Department of the Air Force (DAF) is currently moving very rapidly where human-AI integration is concerned. The research conducted as part of this line of effort aims to enable actionable recommendations about who should initiate human-AI task transfers, the influence of transparency on AI reliance, and specifics about implementing human-AI teaming.

The overriding goal of our team is to optimize task and information sharing between human operators and AI-based automation partners. As such, the RJCA Line of Effort (LOE) is a mix of push, pull, and exploration. It provides data-based recommendations to optimize human-autonomy task sharing and refine AI task coordination. This is critical for all DAF initiatives, because human decision-makers must be kept in the loop while AI is granted responsibility across a variety of applications and systems.

RJCA directly addresses the increasing complexity of human-AI interactions within the ISR realm. Its focus on agentic AI has high relevance for ISR critical projects such as Best of Breed Configurable Adaptive Teammate (BoBCAT), Legion, and Counter-Information, Surveillance, and Reconnaissance (C-ISR). Such 6.3 efforts are key products on AFRL's Integrated Product Teams (primarily Long Range Kill Web and Counter Targeting). RJCA research will help prepare the strategy for future human and AI collaboration to achieve mission success. If successful, the data obtained will enable enhanced control and information sharing between human and AI teammates to achieve synergy and ensure appropriate operator collaboration with AI in real time to support mission execution.

Given the explosion of AI-enabled automation technology across Air Force Speciality Codes (AFSCs), our research investigations explore ways to ensure appropriate compliance and reliance on AI technology in real time. Ultimately, we seek to provide actionable information about trust, transparency, and operator state to our 6.3 transition to optimize human-automation coordination. The following sections will detail progress made on the Closed-Loop Human-Automation Interaction and Multilevel Transparency for Human-Autonomy Interaction projects. ☆

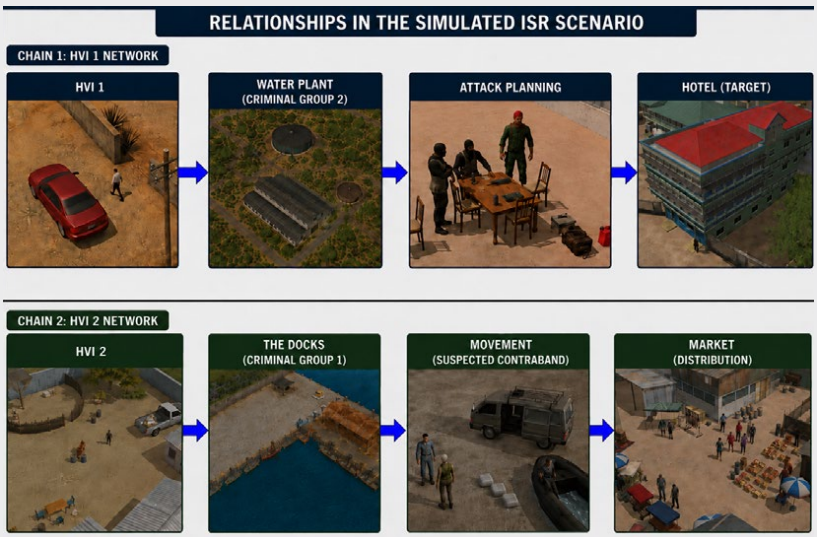
Program Leads

LOE LEAD
Dr. James Bliss

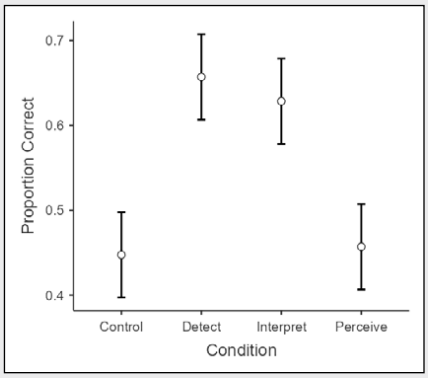
PROGRAM MANAGER
Capt Brody Jones

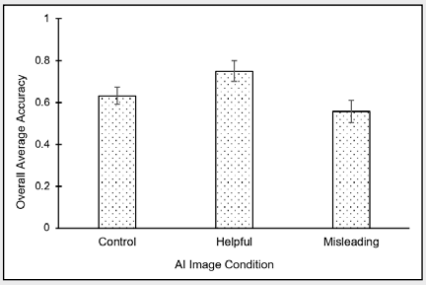
Dr. James Bliss, Rapid Joint Cognitive Awareness LOE Lead and Senior Research Psychologist, 711 HPW/RHWID

Figure 1. Screen shot from the Simulation Platform for Intelligence Collection and Evaluation (SPICE) task.



Graphics and graphs by Anna Maresca (Parallax)





Figures 2, 3. Human Performance Results from SPICE1 (AI Team Role) and SPICE2 (AI Helpfulness) Experiments.

Closed-Loop Human-Automation Interaction (CLHA)

The interplay between human operators and automated systems has remained a critical focus of Air Force researchers. Many Air Force systems feature a role for artificial intelligence/machine learning (AI/ML) coordination [Machine Operator Blending for Integrated Uncrewed Systems (MOBIUS), Joint All Domain Persistent Adaptive Collaborative Technologies (JADPACT), Legion, and Best of Breed Configurable Adaptive Teammate (BoBCAT)]. It is increasingly important that human operators remain cognizant about automated operations and that sharing of responsibilities occurs seamlessly in real time.

Many researchers and task developers have focused their energies on sharing entire tasks between human operators and automation. Given the particular need for human Intelligence, Surveillance, and Reconnaissance (ISR) operators to stay in the loop, perceptually and cognitively, recent efforts have begun to concentrate on ISR-relevant use cases that feature true sharing of tasks between human and automated partners. Successfully achieving such sharing would likely ensure a more effective balance of workload and situation awareness while allowing human operators to demonstrate improved resilience when resuming task responsibility.

The CLHA adaptation effort has been a collaboration between government and industry researchers in RHW and the Air and Space Biosciences Division (RHB) to investigate using physiological data to initiate human-automation task control tradeoffs seamlessly and effectively. Over the last three years, researchers from RHW and RHB have addressed six questions: WHO should initiate task transfer, WHEN should transfer occur relative to mission demands, HOW should transfer be signaled to the operator, WHERE should indications occur on the human operator's screen, WHY should transfer occur based on operator state, and WHAT elements of the task should transfer during a mission.

In FY 2026, research progress occurred in several important areas. The experimentation team developed three testbeds that represent increasing levels of ISR task fidelity. The ISR MATB (Multi-Attribute Task Battery) represents a low-fidelity approach to study human-AI teaming during ISR-relevant workload extremes. VAMPIRE (Visual-Auditory-Motor Performance in Integrated Response Environments) features medium-fidelity cognitive and perceptual task loading as participants respond to signals and identify EEIs (Essential Elements of Information). The highest fidelity testbed, SPICE (Simulation Platform for Intelligence Collection and Evaluation), requires participants to identify threats after viewing intelligence narratives and videos (see Fig. 1). This year, researchers completed online and in-person data collection sessions using VAMPIRE as well as two online sessions using SPICE. To date, the team has individually tested 427 participants to address the questions discussed above.

Results from recent investigations have focused not only on particulars of human-AI control transfer, but the specific cognitive responsibilities that AI should assume and the reactions of operators when the AI offers "helpful" or "misleading" directions (see Figs. 2 and 3).

Additional progress has been made across several areas. RHB researchers have developed a physiology-based model for potential application as a trigger for control transfer to occur. The model relies on machine learning and has demonstrated 92.6% classification accuracy. The CLHA team has worked closely to share testbed development activities and participant testing results with industry partners and 6.3 transition partners.

Future activities for CLHA include archival analysis of existing data, data collection to supplement investigations that have already been completed, and continued collaboration with industry, academic, and government partners. ☆

Dr. James Bliss, Senior Research Psychologist, 711 HPW/RHWID

Dr. Robert Patterson, Senior Research Psychologist, 711 HPW/RHWID

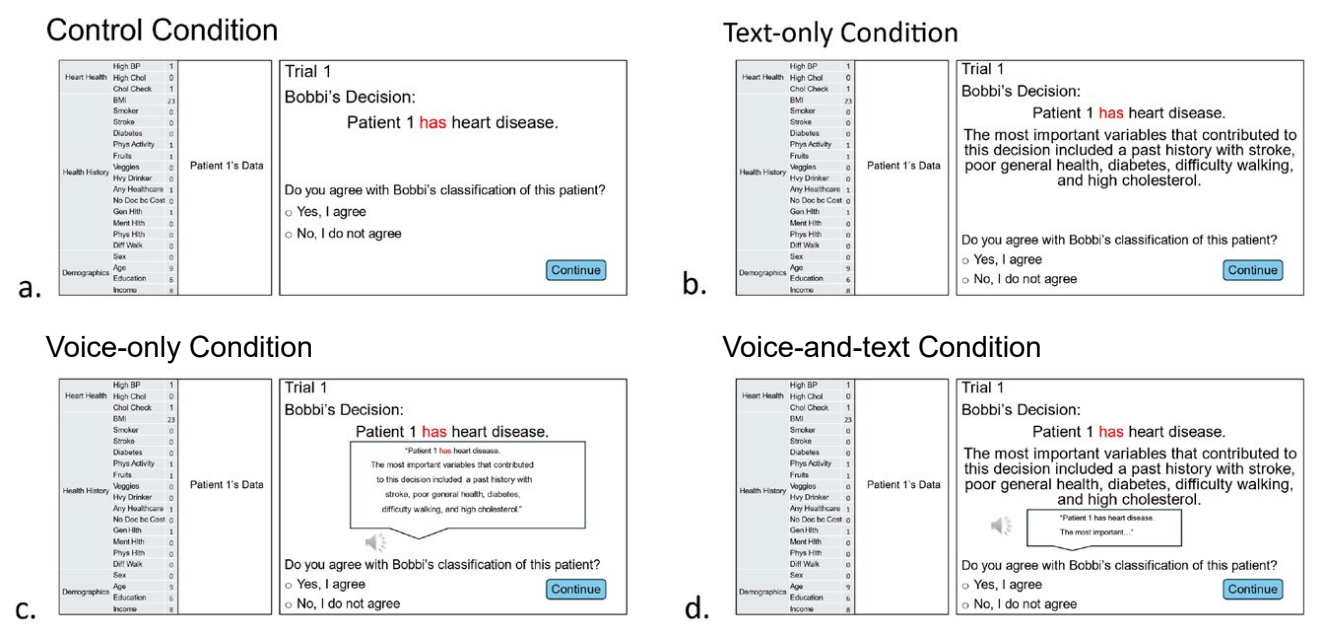


Figure 1. Screenshot Images from the Task of Bobbi's Decision Across Four Explainability Conditions. Charts by 711 HPW/RHWI

Multilevel Transparency for Human-Autonomy Interaction

Transparency is a psychological construct describing the degree to which end-users perceive machine systems as explainable, interpretable, and understandable (Alarcon & Capiola, 2025). System transparency guides end-users' trust toward and reliance on machine systems, which ultimately affects human-machine teaming (HMT) performance outcomes (Chiou & Lee, 2023). Given the opacity in emerging machine systems leveraging artificial intelligence (AI), the USAF (AFDN 25-1, *Artificial Intelligence*) emphasizes the import of system transparency. As transparency in human-autonomy interaction will act as a force multiplier in the Future Fight, research explicating when, why, and for whom transparency is relevant in human-AI interactions becomes increasingly necessary. It is in this vein that the Multilevel Transparency for Human-Autonomy Interaction Project (MTHAI) is identifying ways to promote AI-enabled system transparency through systematic, human-centered experimentation with the goal of facilitating calibrated trust toward and effective use of complex machine systems.

In FY 2026, MTHAI investigated the influence of machine learning (ML)-enabled transparency affordances in decision support systems for non-image-based data classifications. These findings were delivered to 6.3 partners in the Collaborative Technologies branch (RHWT) and submitted for publication in peer-reviewed outlets. Future research iterations investigating large language models (LLM) with varying transparency and capabilities were planned leveraging a custom LLM-enabled testbed despite the elimination of 6.2 Core-funding in FY 2027. MTHAI has successfully capitalized on the advancements of commercial LLMs as well as their own extant human-subjects results, positioning RHW at the cutting edge of human-centered interaction with AI, ML, and LLM-enabled systems.

In addition to 6.2 Core-funded experimentation, MTHAI completed their 3-year 6.1 Air Force Office of Scientific Research (AFOSR)-funded effort identifying foundational transparency guidance for human-AI/ML interactions and methods of promoting eXplainable AI (XAI) for effective human-machine teaming. MTHAI provided critical transparency guidance to cross-Directorate stakeholders concerned with information, sensor, and space domains. MTHAI also continues their international collaboration with NATO partnering nations on human-centered design promoting successful human-robotic swarm interactions, as well as cross-directorate guidance for human-centered transparency considerations in human-swarm interaction through the WARTECH process.

MTHAI deliverables include transparency guidelines for Advanced Technology Development programs aligned with Long Range Kill Web (LRKW) and CT IPTs, as well as inputs for cross-Directorate stakeholders [Information and Spectrum Warfare Directorate (RF), Space Warfare Directorate (RJ), and the Technology Transition Office (RR)]. Some identified interface concepts promoting transparency include a combination of voice and text modalities for effective HMT in data classification contexts and publishing guidelines coupling accurate algorithms with tractable, explainable software for non-image-based data classification. Ongoing work focuses on tailoring explanation features of a decision support assistant ("Bobbi") to end-users and chain-of-thought reasoning to promote transparency in human-LLM interaction (see Fig. 1). Collectively, MTHAI explicates effective human-AI interactions, an emerging human-machine teaming context, to promote the 711 HPW Decision Superiority Operational Concept. ☆

Dr. Gene Alarcon, Senior Research Psychologist, 711 HPW/RHWIM

Dr. August Capiola, Research Psychologist, 711 HPW/RHWIM

Mr. Dexter Johnson, Research Computer Engineer, 711 HPW/RHWIM

Ms. Sasha Willis, Associate Research Psychologist, 711 HPW/RHWIM

READINESS PL

Demonstration of first-ever operational instantiation of a Proficiency-Based Training ecosystem at the 7th Fighter Training Squadron for ACC's Deputy Director of Operations (ACC/A3-2), ACC's Chief Scientist (ACC/ST), and the 1st Fighter Wing's Operations Group Commander (ACC 1 OG/CC).

GRILL

Aircraft Maintenance Officer Course Simulation: Developed and delivered a new maintenance-training simulator for AETC that strengthens flightline coordination and management skills during capstone training for more than 300 Airmen each year.

WMI PL

Completion of Multi-Decision Advantage Sprint for Human-Machine Teaming (MASH) event, led by the JADPACT team, demonstrated successful integration of a multi-vendor solution for battle management interface, incorporating data collection and operator feedback.

Cooperative Research and Development Agreement between MOBIUS and industry partners established to facilitate collaborative HMI development for CCA management.

IAS PL

Presented Kraken: Intelligence, Surveillance, and Reconnaissance feasibility and optimization engine to the Air and Space Force Association conference with an audience including over 22 stars worth of General Officers.

Legion presented capabilities at Combat Air Forces Weapons and Tactics Conference establishing them as a leader in Joint All Domain Command and Control, resulting in fruitful relationships with many operational partners such as the AFCENT Battle Lab.

WIT CTC

CCA-ASR: Utilized advancements in 6.2 research in both speech recognition and text-to-speech to successfully tune and evaluate speech recognition for a specific military domain out-performing commercial systems. The resulting model was transitioned to the DARPA VENOM program and the results of this work will inform the CCA PoR with future decisions on CCA control interfaces.

BoBCAT-RSO: Transitioned agent generated sensor based algorithm (SBA) development to AFLCMC/ROD providing results in minutes compared to the manual process of days/weeks increasing the throughput and accuracy of SBA generation significantly.

SAVAGE: Transitioned custom Text-To-Speech interface to the 645th AESS in support of content generation for cryptolinguist training. The ability to create custom speech and adjust for pronunciation and emphasis fulfills a long requested requirement from the program office.





RHW SUCCESSES FY26





HLC CTC

Delivered updated pilot flight manual with flight duration guidelines and Standard Operating Procedures for F-35 engine operations to F-35 Joint Program Office to reduce risk of hearing related injuries and enhance CONOPS.

Transitioned cognitive fatigue modeling application program interface (API) to industry partners for testing and evaluation to enhance warfighter fatigue risk management tools.

Developed multimedia influence data pipeline for playbook capability in a cognitive warfare intelligent automation system to enhance Information Operations.

Demonstrated FLARE's real-time operator performance tracking with Electronic Warfare operators during the Training Integration and Experimentation Research (TIER) live exercise in Playas, New Mexico, identifying research priorities to significantly accelerate after-action review workflows.

RHW and the Air and Space Biosciences Division (RHB) modernized USAF en route care training by deploying the Team Dynamics Measurement System, gathering data from ~1,500 trainees to deliver AI-driven performance metrics that actively shape curriculum, staffing, and future operations.

ACRONYMS:

ACC = Air Combat Command | AESS = Aeronautical Systems Group
AETC = Air Education & Training Command | AFCENT = US Air Forces Central
AFLCMC = Air Force Life Cycle Management Center
ASR = Automatic Speech Recognition
BoBCAT = Best of Breed Configurable Adaptive Teammate
CCA = Collaborative Combat Aircraft | CONOPS = Concept of Operations
HMI = Human-Machine Interface | DARPA VENOM = Defense Advanced Research Projects Agency Viper Experimentation & Next-gen Operations Model
FLARE = Framework for Language Analysis & Resilience Enhancement
JADPACT = Joint All-Domain Persistent with Adaptive Collaborative Technologies
MOBIUS = Machine-Operator Blending for Integrated Uncrewed Systems
PoR = Program of Record | RSO = Rapid Sustainment Office
SAVAGE = Synthetic Agent Voice Analysis Generation and Exploitation
SBA = Sensor Based Algorithms

EST. 2001

To view more issues, visit:
www.afresearchlab.com/FightsOn

Published since 2001, Fight's ON! continues to serve as the Division publication for our partners and features innovative science and technology that is accelerating and revolutionizing readiness.

Fight's ON! Point-of-Contact
Patricia D. Wood, 711 HPW/RHW
patricia.wood.2@us.af.mil

Cleared / Case # AFRL-2026-2022

711 HPW/RHW Core Research Areas:
- Digital Models of Cognition
- Distributed Teaming & Communication
- Human-Machine Interactions
- Learning & Operational Training

Distribution Statement A / Approved for public release; distribution is unlimited.