



**Artificial Intelligence, Automation, and Augmentation Program**

**ARLIS Technical Point-of-Contact:**

Craig Lawrence, Ph.D.  
 Director, Systems Research  
 Applied Research Laboratory for Intelligence and Security  
 301.226.8808; [clawren4@umd.edu](mailto:clawren4@umd.edu)

**ARLIS is closing the gap in our ability to test and evaluate the operational impact of emerging technologies**

A great deal of effort is being spent developing artificial intelligence (AI)-based and automated tools, as well as human augmentation technologies, to support warfighters and intelligence analysts. Comparatively little effort is being made to develop the methodologies and supporting technologies to assess the performance of these tools in an operational mission context. This gap results in research and development investments not transitioning into operational use, or potentially worse, ineffective technologies transitioning into operational use.



**ARLIS is developing a prototype testbed and methodologies to evaluate AI, automation, and human augmentation (AAA) technologies**

To meet this need, ARLIS is developing a prototype testbed and methodologies to evaluate AI, automation, and human augmentation (AAA) technologies – characterizing the overall quality of the support they provide, and identifying how they might be best employed, or where they might be harmful. To ensure that our work is grounded in mission needs, ARLIS is working with mission partners who are providing their operational insights (and access to their operators, tools, and data where available) to validate our developing approaches, including the Joint Artificial Intelligence Center (JAIC), the special operations community, and NGA-Research.

**ARLIS incorporates accuracy into a greater model of mission-centric performance**

ARLIS has identified two key steps that are necessary to answer the questions of operational impact. First, it is critical to identify and use the correct metrics. The field of AAA practitioners tends to leverage accuracy metrics as their preferred method for describing performance. Accuracy is a reasonable way of comparing relative performance among a set of technologies but does not directly map to operational impact. Accuracy needs to be incorporated into a greater model of mission-centric performance, and we need to clearly convey what subset of mission conditions has been tested.

**ARLIS builds assurance cases combining results across traditionally stovepiped test domains**

We recognize that testing the full space of operational conditions – especially considering user-in-the-loop operations, is prohibitive. Instead, ARLIS is building assurance cases by combining results across test domains that are traditionally stovepiped: operational test in the field, operational data logging, formal experimentation in controlled laboratory conditions (with both expert and non-expert users), and advanced simulation of user decision making. Each of these approaches provides a different balance of quality and quantity: that is, the goodness of the results, versus the opportunity to run a large variety of tests under varying conditions. The goal is to intelligently combine the findings across these evaluations to make optimal use of the limited opportunity for operational test and evaluation in the field.