SERC Talks: "Progress in Test and Evaluation of AI-enabled Systems in the DoD"

October 29, 2021 | 1:00 PM ET

Dr. Yevgeniya "Jane" Pinelis
Chief, Test, Evaluation, and Assessment,
Department of Defense Joint Artificial Intelligence Center (JAIC)

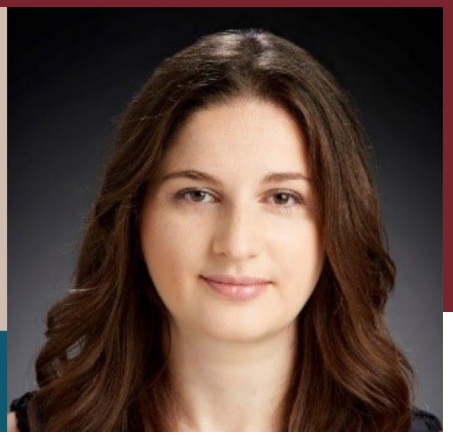
TEST & EVALUATION

- ❑ Today's session will be recorded.
- ❑ An archive of today's talk will be available at: www.sercuarc.org/serc-talks/ as well as on the SERC YouTube channel.
- ❑ Use the Q&A box to queue questions, reserving the chat box for comments, and questions will be answered during the last 5-10 minutes of the session.
- ❑ If you are connected via the dial-in information only, please email questions or comments to SERCtalks@stevens.edu.
- ❑ Any issues? Use the chat feature for any technical difficulties or other comments, or email SERCtalks@stevens.edu.

SYSTEMS ENGINEERING RESEARCH CENTER

SERC Talks: "Progress in Test and Evaluation of AI-enabled Systems in the DoD"

SERC TALKS



Dr. Yevgeniya "Jane" Pinelis
Chief, Test, Evaluation, and Assessment,
Department of Defense
Joint Artificial Intelligence Center (JAIC)

TEST & EVALUATION

Dr. Laura Freeman, SERC Research Council Member, Director of the Hume Center's Intelligent Systems Lab and Assistant Dean for Research in the College of Science, and Research Associate Professor in the Department of Statistics at Virginia Tech

SYSTEMS ENGINEERING RESEARCH CENTER

DISCLAIMER

The Systems Engineering Research Center (SERC) is a federally funded University Affiliated Research Center managed by Stevens Institute of Technology.

Any views, opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense, OUSD (R&E), nor the SERC.

No Warranty. This SERC - Stevens Institute of Technology Material is furnished on an “as-is” basis. SERC and Stevens Institute of Technology makes no warranties of any kind, either expressed or implied, as to any matter including, but not limited to, warranty of fitness for purpose or merchantability, exclusivity, or results obtained from use of the material. SERC and Stevens Institute of Technology does not make any warranty of any kind with respect to freedom from patent, trademark, or copyright infringement.

This material has been approved for public release and unlimited distribution.

Test & Evaluation of AI Capabilities: 2021 Status Update

Dr. Jane Pinelis, Chief, T&E – Joint Artificial Intelligence Center

October 29, 2021

T&E is an opportunity for asymmetric advantage in the AI arena.

DoD does not need to be fastest—we should develop **fast enough** and **better**.

DoD must update its T&E process for AI
But we have to come up with our own solutions

Tasks are dynamic and poorly constrained, success is hard to define/measure, and failures are catastrophic.

Can test only fraction of AI's operational space, but generalizing test results is difficult to impossible.

DoD must integrate adaptive T&E process across cradle-to-grave system lifecycle

AI T&E is coalescing, but work remains

Current integration efforts fall under three umbrellas

Research

Developing methods that will allow us to measure, model, and analyze AI system test results.

Frameworks

Guidance that will empower non-experts to create reasonable first passes at different aspects of test plans for these extremely complex systems.

Resources

Instrumentation, ranges, test beds, data infrastructure, and automated analyses that will enable effective T&E of AI.

RESEARCH: We are advancing current T&E methods and inventing new methods

HSI and HMT	*ARLIS, AFRL, IDA, JAIC, DOT&E, MITRE*
Sequential Test Design	*IDA, DOT&E, JHUAPL, VTIC*
Learning Systems	*AFRL, DARPA*
XAI	*ARL, DARPA, NRL*
RAI	*DARPA, JAIC*
Adversarial/Red Teaming	*DARPA, MITRE*

FRAMEWORKS: We are building tools to support working level testers face the challenge of AI T&E

AIES T&E Framework	*JAIC, DoD AI ESG WG*
Assurance Cases	*IDA, STAT COE, DSTL*
System Integration	*JAIC*
Human System Integration	*JAIC, CMU, Sandia, PNNL*
Operational & "Tactical" Testing	*JAIC*
Responsible AI T&E	*JAIC*
Trustworthy Autonomy – System Effectiveness	*IDA*

JAIC AI Testing and Evaluation Process

Integrity Testing
Test performance in normal situation

Confidence Assessment
Assess Calibration & Uncertainty Awareness

Robustness
Test performance on natural perturbation and transformation

Resilience
Perform adversarial testing

Tailorable testing approach where each step is scaled to mission requirements

System Integration Testing

Functionality	• The ability of the system to do the work for which it was intended
Reliability	• The probability that a system performs correctly during a specific time duration.
Interoperability	• The ability of computerized systems to connect and communicate with one another readily, even if they were developed by different manufacturers in different industries.
Compatibility	• Is a technique by which 2 or more application interact in the same environment.
Security	• Testing the systems vulnerability and weakness

RESOURCES: We are developing needed tools, training, and infrastructure to execute T&E

Test Harness	*JAIC, JHUAPL, AFRL, ARL, TRMC*
T&E BPA	*JAIC, ACC-RI, IN3*
RAPT	*JHUAPL*
Test Data capability	*TRMC, JAIC, VA Tech*
T&E Factory	*JAIC, TRMC*
T&E of Autonomy	*DAU, TRMC*

T&E BPA – Overview

- **To make T&E AI-ready, the acquisition package must:**
  - **Support JCF as central hub for AI development and AI T&E**
    - Central platform for tailored, extensible T&E, curation of training & validation data, and identification of best metrics to characterize performance
    - Accessible by DoD components
  - **Address entire T&E lifecycle**
    - To include AI Test Tools, Human-System Integration (HSI) Testing, and Operational Testing of AI-enabled systems
    - Evaluate algorithms for characteristics of responsible, trusted AI to include transparency and explainability criteria
    - Individual uses will require flexibility in selecting a subset of the BPA categories
  - **Support T&E and IV&V of current and emerging AI capabilities**
    - Algorithms must undergo T&E to determine suitability for use in operations
    - Unique challenges in AI operationalization
    - Wide breadth and forward looking

BPA Service Categories

- Test Technology and Tools
- Data Set Development/Curation
- Test Harness Development
- Model Output Analysis
- Test Planning, Documentation, and Reporting
- Testing Services
- Targeted T&E and IV&V
- Algorithm T&E
- System T&E
- Operational T&E
- Human Factors/User Acceptance T&E
- Model & Simulation
- Adversarial AI Testing
- Emergent Behavior Testing
- Project Management

Categories address unique AI test tools, technology, & services

JAIC HSI AI Framework

O

O

"I need to understand and predict the situation. I need what I need to know, when I need to know it, in a way that I understand."

D

"I need to make be able to make good decisions about where and how to use this system."

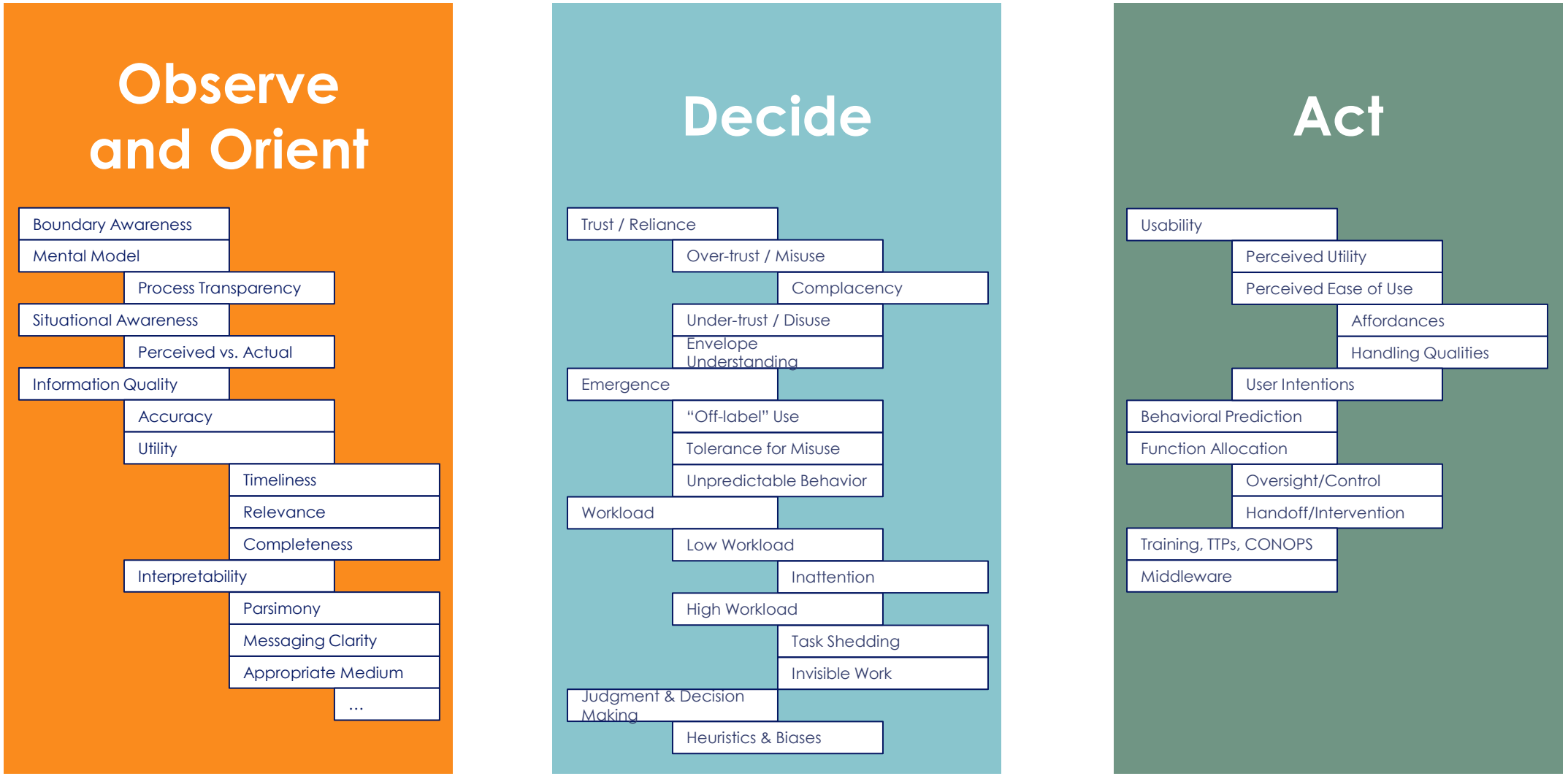
A

"I need to be able to get the system to do what I need and intend it to do."

DoD Ethical AI Principles

1. Responsible
2. Equitable
3. Traceable
4. Reliable
5. Governable

JAIC HSI-AI Framework



JAIC OT Framework

Tactical Testing

- Smaller, but more frequent tests building towards IOT&E
- Multiple contexts and environments
- Variety of threats
- Stress individual aspects of the mission

Decision making is performance

- 3 types of decisions
  - Perceptual (“What?”)
  - Procedural (“How?”)
  - Executive (Goals setting)
- Influences how we test

Robust failure modes ensure safe testing

- Cannot test for everything
- Known unknown and unknown unknown failures will occur
- Need failures to be “graceful”

Scaling to enterprise will require policy and coordination

- **DOT&E and JAIC are building an AI&A roadmap:**

1. Integrate T&E across the cradle-to-grave capability lifecycle.
2. Improve and develop new T&E test methods
3. Invest in infrastructure, data tools, and other common solutions.
4. Institute systemic processes and common architectures.
5. Build a workforce that is mission-ready in both AI and T&E.
6. Establish long-term collaborations among DOD, other government entities, industry, academia, and allies.
7. Leverage research, frameworks, and collaborations to address unsolved AI&A T&E challenges (e.g. RAI, Cyber, M&S).

Final thoughts

- Rigorous and timely T&E capability is a strategic advantage
- Science, policy, and infrastructure for AIES need to keep pace with technology development and modernization
- We have made tremendous progress, especially in changing culture
- Much remains to be done!
- There is a very special role for FFRDC and UARC collaborations

Questions?

Connect with us:
AI.mil
Twitter: @DoDJAIC
LinkedIn: DoD Joint Artificial Intelligence Center

SYSTEMS ENGINEERING RESEARCH CENTER

CONTACT US:

serc@sercuarc.org

www.sercuarc.org/contact-us/

SERC TALKS

"Test and Evaluation" Series

December 1, 2021 | 1:00 PM ET

Dr. Sandra Hobson

Deputy Director for Strategic Initiatives, Policy and Emerging Technologies,

Office of the Director, Operational Test and Evaluation,

Office of the Secretary of Defense

CONTACT

Webinar Coordinator: Ms. Mimi Marcus, Stevens Institute of Technology – mmarcus@stevens.edu

Please visit the SERC Talks page to register and for more information and updates.

REGISTER NOW: SERC SPONSOR RESEARCH REVIEW 2021

All times listed in Eastern Time

- TUESDAY: November 2 | IN-PERSON (Washington, DC) / VIRTUAL
  - Learn about the new Acquisition Innovation Research Center (AIRC) and explore Initial AIRC Research
  - Join us for a Keynote by Mr. Christopher C. O'Donnell, Performing the Duties of the Assistant Secretary of Defense for Acquisition
  - Panel on Acquisition Innovation
- WEDNESDAY: November 3 | VIRTUAL
  - Join us for a Keynote by Mr. Maynard Holliday, Director of Defense Research and Engineering for Modernization
  - Two Panels: Modernization Priorities, and Systems Engineering Modernization
  - Explore the latest SERC research in Digital Engineering, AI & Autonomy, Velocity, Security and Human Capital Development
- THURSDAY: November 4 | VIRTUAL
  - NEW! Free tutorials: Digital Engineering & Security Engineering
  - Explore the latest SE doctoral research

- **REGISTER HERE: https://sercuarc.org/research-reviews/2021-serc-annual-research-review/#registration**

SYSTEMS ENGINEERING RESEARCH CENTER

SERC TALKS

THANK YOU FOR JOINING US!

Please check back on the SERC website for today's recording and future SERC Talks information.

Subscribe and follow SERC on our social channels