

# Validation for AI and Autonomous Systems

**WRT 1011 Sponsor: OUSD(R&E)**

**By**

**Dr. Valerie Sitterle, Dr. Paul Collopy and Ms. Jennifer Petrillo**

**11<sup>th</sup> Annual SERC Sponsor Research Review**

**November 19, 2019**

**FHI 360 CONFERENCE CENTER**

**1825 Connecticut Avenue NW, 8<sup>th</sup> Floor**

**Washington, DC 20009**

**[www.sercuarc.org](http://www.sercuarc.org)**



**Trained in field; may have *a priori* training**  
**Will establish and evolve algorithm and/or**  
**parameters based on field data**

**Trains in field; may have *a priori* training**  
**No changes to algorithm**  
**Will establish and evolve parameters**  
**based on field data**

**Trained then fielded**  
**Will change algorithm and/or update**  
**parameters based on new data**

**Trained then fielded**  
**No field changes to algorithm**  
**Will update parameters based on new data**

**Trained then fielded 'as is'**  
**No field changes to algorithm or parameters**

**General AI**  
**(not there yet...)**

**Learning**  
**Systems: AI/ML**

**Narrow ML/AI:**  
**Mutable**

**Narrow ML/AI:**  
**Semi-Mutable**

**Narrow ML/AI:**  
**Fixed**

**Human**

- Handles novel situations, evolving environments
- Performance may be inconsistent
- Exhibits only moderate repeatability
- Behavior may be unpredictable

**Autonomous**

- Handles only constrained situations, defined environment, and narrow task(s)
- Performance typically consistent
- Performance repeatable for designed situations
- Behavior predictable

**Automated**

- **Narrow AI – Task specific & Fixed SW**
  - Same engineering brittleness – and engineering considerations – as other automated systems
- **The impact of MUMT**
- **Increasing complexity with increasing autonomy**
  - System capabilities expanded to handle more operational situations
  - Reduced **Understandability** & **Predictability** of system performance
  - Increased **Vulnerabilities** via expanded attack surfaces, especially if MUMT required

- **Learning Systems ML/AI**
  - **When to field?**
    - When is training enough? Before fielding?  
When learning in the field?
    - When is “good enough” trusted for field operations?
  - **Data integrity?**
    - Effort to validate, cleanse for training
    - Can’t do this well in field
  - **How much learning?**
    - If system adjusts in field, do we really want this to be automated?
  - **Asset transferability?**
    - Training in one theater not often transferable to another: *Here ≠ There*
  - **What happened to the Digital Twin?**





# How to test a fully autonomous system?



# Toulmin Model of Argumentation

