



ARMD Transformative Aeronautics Concepts Program (TACP) Convergent Aeronautics Solutions (CAS) Execution Project



Autonomy Teaming & TRAjectories for Complex Trusted Operational Reliability

ATTRACTOR

ARMD Technical Seminar

14 May April 2020

Co-PIs: Natalia Alexandrov (LaRC), Danette Allen (LaRC)





Origins: Two 2017 Proposals

HOLII GRAILLE: Enabling Teaming and Measuring Trust in Autonomy (PI/co-PI: BDA/NMA)

Idea: Engender trust in human-machine teaming via:

- 1) natural human-machine interaction (speech, gestures, mixed-reality) and
- 2) reliable behavior in time-critical and safety-critical situations.



System for Identifying Reward Functions for Time-Critical Human-Machine Decision Making (PI/co-PI: NMA/BDA)

Idea: Treat every action in cyber-physical-human (CPH) ops as a multiobjective optimization problem under uncertainty. Engender trust and trustworthiness via correct objectives and shared decision-making models.

CAS encouraged **merging** of proposals related to Autonomy. These two proposals served as a starting point for the development of the ATTRACTOR proposal.





ATTRACTOR
TEAM

FIN

05/24/2020

FY20

FY19

FY18

ARMD Technical Seminar

NASA + ARMD
2019 WINNERS



ATTRACTOR Team



Autonomy Teaming & TRAjectories for Complex Trusted Operational Reliability

LaRC

Edward Adcock
Bryan Barrows
Jeremy Castagno¹
Meghan Chandarana^{1,6}
John Cooper
Chester Dolph
Brian Duvall²
James Ecker
James Farrington
Alwyn Goodloe
Lynn Hadley
Zachary Johns²
Benjamin Kelley
Lisa Le Vie
Kyle McQuarry
Andrew Miloslavsky

Paul Miner
Daniel Moerder
James Neilan
James Plant
Paul Rothhaar
Ivan Perez²
Javier Puig-Navarro³
Loc Tran
Matthew Vaughan
Ralph Williams
UAS Operations Office

Jon How⁴
Nick Roy⁴
Sebastian Scherer⁶
Matthias Scheutz⁵

ARC

Adrian Agogino
Kamalika Das
Dimitra Giannakopoulou
David Iverson
Ritchie Lee
Nikunj Oza
Donald Soloway
Shawn Wolfe

AFRC

Manuel Castro
Jason Holland
Otto Schnarr
Erik Coltey¹

GRC

Ricaurte Chock
Kristen Hauser
Paul Nelson



- [1] Interns
- [2] National Institute of Aerospace
- [3] University of Illinois Urbana Champaign
- [4] Massachusetts Institute of Technology (MIT)
- [5] Tufts University
- [6] Carnegie Mellon University (CMU)



The Problem

AAM/UAM – target environments



3D!

Unprecedented complexity



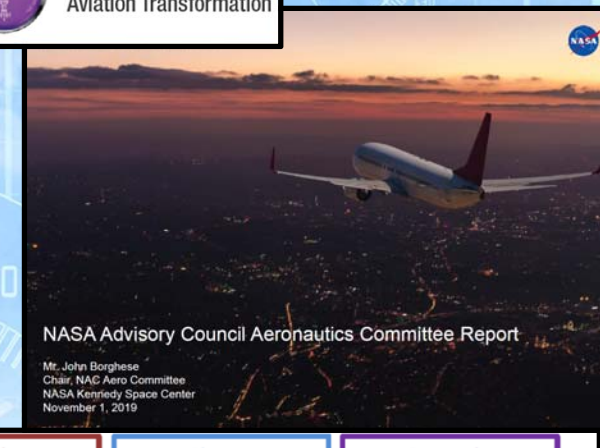
How to achieve **trustworthiness and justified trust** in autonomous cyber-physical-human (CPH) systems in **safety-critical, time-critical, and certification-dependent** environments?



ATTRACTOR Relevance



- ARMD Strategic Thrusts & Outcome Addressed: 1, 5, 6
 - NASA Advisory Council Aeronautics Committee Report (November 2019)
 - “NASA’s aeronautics domain expertise bridges the gap between intelligent systems and civil aviation
 - Applies emerging machine intelligence technologies to aviation applications
 - Applies domain expertise in integrated human-machine operations, airspace management, and aviation safety.”
 - McKinsey and Booz Allen Hamilton UAM market studies:
 - “For UAM to be viable, it is necessary to address the technical, physical, operational, and integration challenges of a highly interdependent system-of-systems.”
 - “Significant legal/regulatory, certification, public perception, infrastructure, and weather constraints exist, which reduce market potential in the near term for UAM.”
- ATTRACTOR addresses technical development requirements outlined in all use cases:
- Autonomous Flight Technology
 - Detect and Avoid
 - GPS-denied technology
- ATTRACTOR is building foundations for developing trustworthiness and (justified) trust in autonomous systems – directly supportive of AAM, UAM and all future transportation models, as well as space exploration environments, that rely on autonomous systems.
 - UML-4: Medium Density and Complexity Operations with Collaborative and Responsible Automated Systems
 - UML-5: High Density & Complexity Operations w/ Highly-Integrated Automated Networks
 - UML-6: Ubiquitous UAM Operations with System-Wide Automated Optimization



	Vehicles	Airspace	Community
UML-1	• Late-Stage Certification Testing and Operational Demonstrations in Limited Environments		
UML-2	• Low Density and Complexity Commercial Operations with Assistive Automation		
UML-3	• Low Density, Medium Complexity Operations with Comprehensive Safety Assurance Automation		
UML-4	• Medium Density and Complexity Operations with Collaborative and Responsible Automated Systems		
UML-5	• High Density and Complexity Operations with Highly-Integrated Automated Networks		
UML-6	• Ubiquitous UAM Operations with System-Wide Automated Optimization		



Trustworthiness $\Leftrightarrow$ Justified Trust

Multi-agent System for
Mission Planning and Execution

Trustworthiness:

- Attribute of cyber-physical-human (CPH)
- Assurance that CPH does what is required and not what is prohibited
- Necessary to deploy CPH in safety-critical environments

+

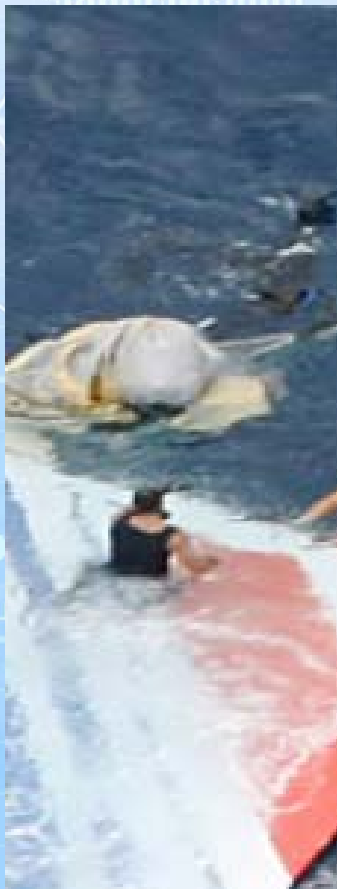
Trust:

- Attribute of the participants/ users/observers/ controllers of the CPH
- Measure of readiness to rely on another system participant for decision/advice/ action

- Trustworthiness should be a prerequisite to trust
- Unjustified mistrust reduces the system's trustworthiness



CPH Trust/Trustworthiness Breakdowns





The ATTRACTOR Challenge

Build a basis for certification of autonomous systems via establishing

- (1) metrics for trustworthiness and trust
- (2) in multi-agent team interactions,
- (3) using AI explainability and
- (4) persistent modeling and simulation, in the context of
- (5) mission planning and execution, with
- (6) analyzable trajectories.

Multidisciplinary Components of
Complex Multi-agent Systems

Measurable trustworthiness
of fundamental functional
components

Context:
Trajectories
and DRM



Sound multi-objective
decision-making

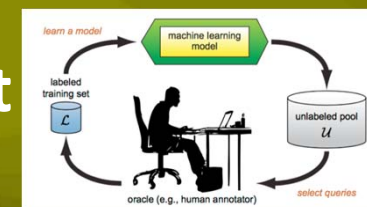
Heterogeneous multi-agent (including human-machine)
teaming is essential

Teaming



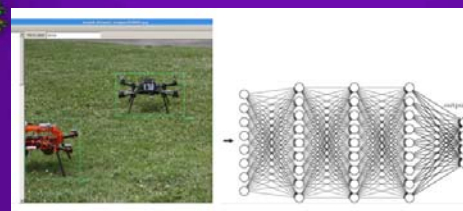
Efficient two-way
learning/ training loop.

Learning &
Justifiable Trust



Good Outcomes:
Basis for Trust

Explainability



These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs.

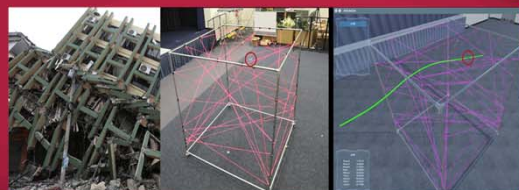


The ATTRACTOR Challenge:

Multidisciplinary Component of
Complex Multi-agent Systems

Measurable trustworthiness
of fundamental functional
components

Context:
Trajectories
and DRM



Heterogeneous multi-agent (including human-machine)
teaming is essential

Teaming



Efficient two-way
learning/ training loop.

**Learning &
Justifiable Trust**



**Good Outcomes:
Basis for Trust**



Sound multi-objective
decision-making

Explainability



These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs.



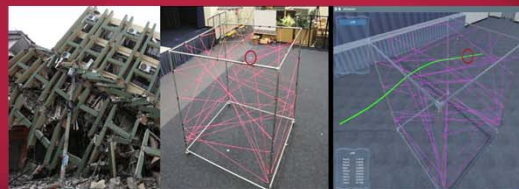
The ATTRACTOR Challenge: Design Reference Mission (DRM)

Multidisciplinary Component of
Complex Multi-agent Systems

Measurable trustworthiness
of fundamental functional
components

Context:

Trajectory
and DRM



DRM for Scope

Heterogeneous multi-agent (including human-machine)
teaming is essential

Teaming



Efficient two-way
learning/ training loop.

Learning &
Justifiable Trust



Explainability



These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs.

Sound multi-objective
decision-making



SAR Mission



Human or Machine Defines Initial Flight Paths



Natural H-M Interaction Facilitates Trust

Generate Mission



Reliable Trajectories



Simulate Mission



Adjust and Approve Mission



Execute Mission

Risk-Aware Execution

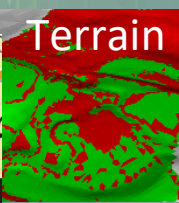
Manage Real-time Risks



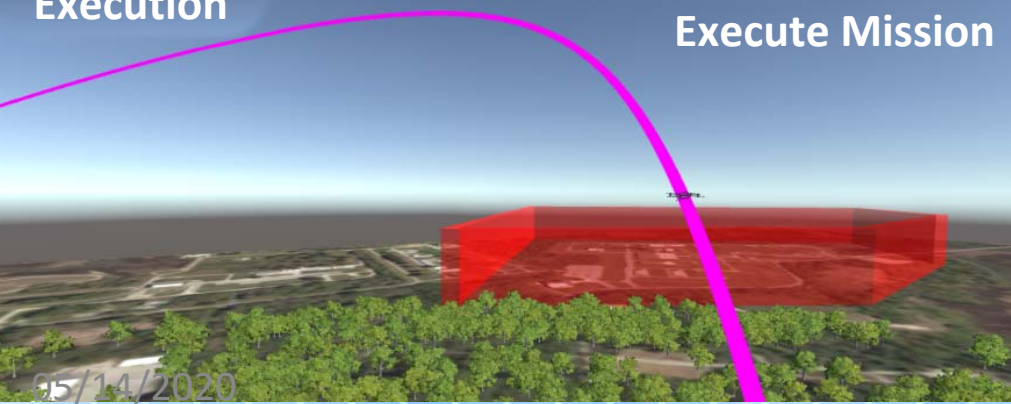
Vehicles



People



Terrain

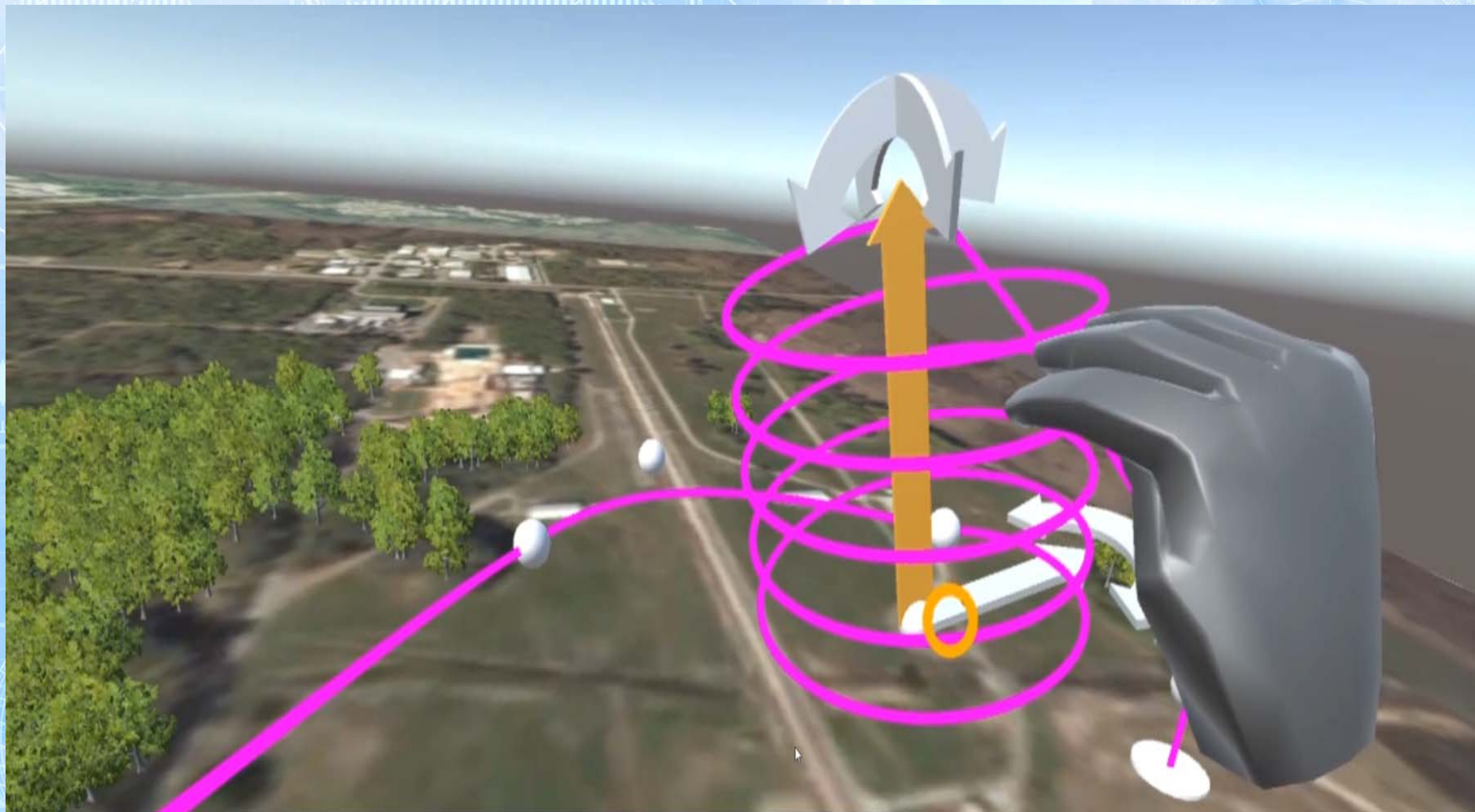


05/14/2020

DRM Storyboard for CPH Context



ATTRACTOR Concept





Multi-Asset SA(A)R Mission



Integrated demo to be conducted if/when we are back on center.
This demo was/is impacted by the furlough and COVID.



AUTONOMOUS
MIXED-REALITY
MULTI-MODAL
MULTI-AGENT
SEARCH & RESCUE



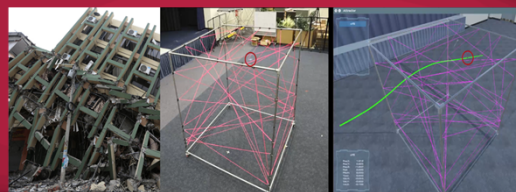
The ATTRACTOR Challenge: Search and Rescue Under The Canopy

Multidisciplinary Component of
Complex Multi-agent Systems

Measurable trustworthiness
of fundamental functional
components

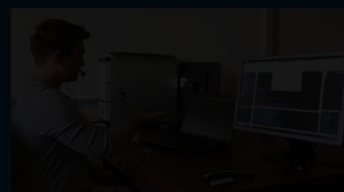
Context:

Trajectory
and DRM



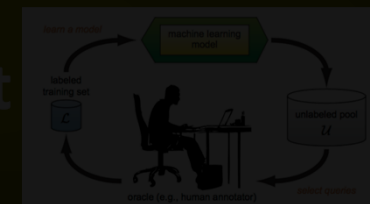
Heterogeneous multi-agent (including human-machine)
teaming is essential

Teaming



Efficient two-way
learning/ training loop.

Learning &
Justifiable Trust



Good Outcomes:
Basis for Trust



Sound multi-objective
decision-making

Explainability



These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs.



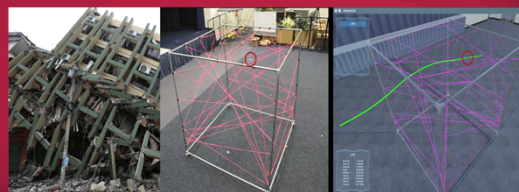
The ATTRACTOR Challenge: Search and Rescue Under The Canopy

Multidisciplinary Component of
Complex Multi-agent Systems

Measurable trustworthiness
of fundamental functional
components

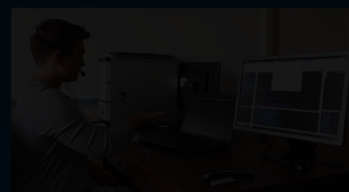
Context:

SARUC



Heterogeneous multi-agent (including human-machine)
teaming is essential

Teaming



Efficient two-way
learning/ training loop.

Good Outcomes:
Basis for Trust

Expl



Sound multi-objective
decision-making

propellers. They have four legs.



SARUC



Search and Rescue Under The Canopy

- Partnership between MIT and LaRC - originated during **CAS DELIVER**
- Transitioned to ATTRACTOR as an instantiation of our DRM
- National press about SARUC flights at LaRC

- <http://news.mit.edu/2018/fleets-drones-help-searches-lost-hikers-1102>
- <https://www.bostonherald.com/2018/11/03/mit-nasa-develop-tech-that-could-be-used-to-find-missing-hikers-or-on-mars-moon/>
- <https://nypost.com/2018/11/05/drone-fleets-could-help-find-missing-hikers/>
- <https://newatlas.com/mit-nasa-search-rescue-drone-fleet/57082/>
- <https://dronebelow.com/2018/11/05/search-and-rescue-under-forest-canopy-using-multiple-drones/>
- <https://www.redbull.com/int-en/mit-autonomous-drone-seeking-lost-hikers>
- <https://www.gpsworld.com/in-mit-project-drones-map-without-gps/>
- <https://www.digitaltrends.com/cool-tech/drones-find-lost-hikers/>
- <https://www.foxnews.com/tech/drone-fleet-could-help-find-lost-hikers-mit-researchers-say>

- AUVSI XPONENTIAL Humanitarian Award 2019

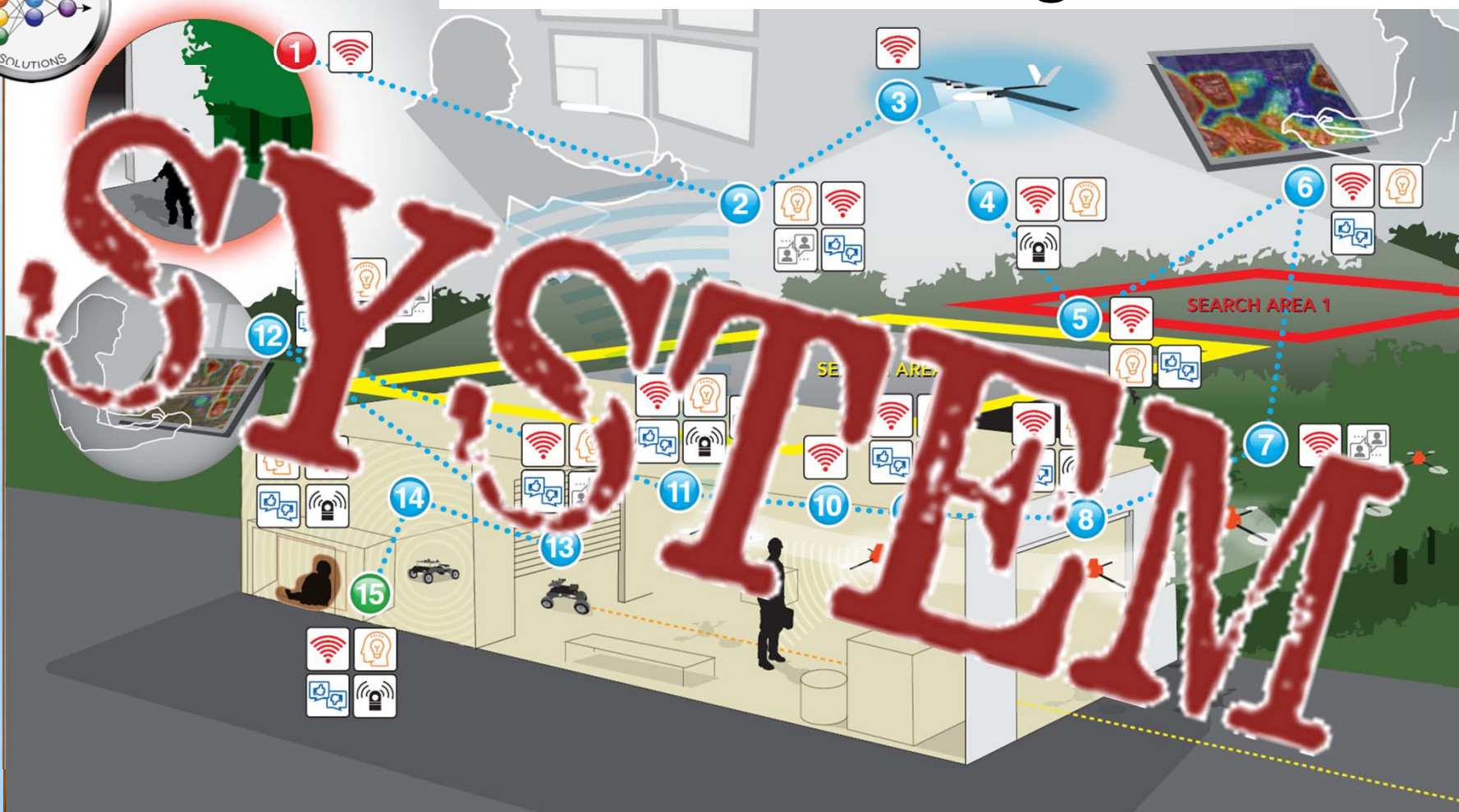




SARUC



ATTRACTOR Integrated Demo



- 1 A person is reported missing.
- 2 First responders identify search area(s) using last known location and missing person data. (age, mobility, clothing, etc.)
- 3 VTOL UAV takes off.
- 4 VTOL UAV captures high altitude video of last known location and surrounding area(s).
- 5 Mission operator identifies a wooded area and a nearby building as areas of interest for an in-situ search.
- 6 An operator uses a touch and speech interface to request a search mission for the exterior and interior of the building of interest with identified risk areas.
- 7 Several UAVs enter into the woods to map the area to search for the missing person while additional UAVs search the exterior of the building and monitor the exits.
- 8 A single UAV enters the building interior through a partially-open bay door.
- 9 The interior drone navigates through a room(s) that is cluttered with equipment and/or furniture.
- 10 While moving through the room, the UAV observes a person, but identifies they are not the missing person of interest and continues its search.
- 11 The interior drone alerts the mission operator that it is too large to proceed through a small opening inside the building.
- 12 The mission operator instructs the rover to search beyond the partially-open bay door.
- 13 The rover enters the building and searches beyond the partially-opened bay door.
- 14 The rover searches the area using its onboard sensors (LIDAR, camera, etc.) and locates the missing person under a desk.
- 15 The rover signals that the lost child has been found and all participants stand down.



Information communication



Decision making



Authority to make decisions



Sensors, e.g., lidar



Human team (first responders) planning/info gathering





Capability Integration into DRM Demo

- Search and Rescue DRM Integrates:
 - Multi-vehicle operation (Trajectory Generation and Intelligent Swarms)
 - Own-ship knowledge (Anomaly Detection, State Estimation)
 - Sensing and perception (Sensors, Object Detection, XAI)
 - User Interface (IM, AI – VAE, Neuro-visualization)
 - Navigation (Verification, Trajectory Stress Testing)
 - Simulation (B.A.M.)
- and requires a unique Modeling and Simulation Environment:
 - Sim-to-Flight:
 - Same software can be used in the simulation environment and deployment of real vehicles
 - A.E.O.N. – vehicle command and control interface
 - LVCDE:
 - Sensor data can be supplied to simulation environment. Or sensor data can be generated in simulation environment to real vehicles. Or both.
 - Unity-based Simulation provides physics engine
 - Plug-n-Play: Publish/Subscribe framework using DDS

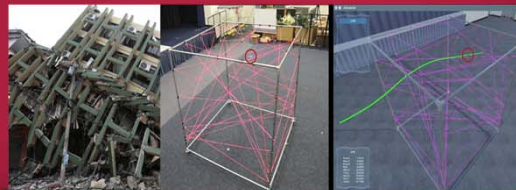


The ATTRACTOR Challenge: Modeling & Simulation (ModSim)

Multidisciplinary Component of
Complex Multi-agent Systems

Measurable trustworthiness
of fundamental functional
components

Context:
Trajectories
and DRM



Heterogeneous multi-agent (including human-machine)
teaming is essential

Teaming



Efficient two-way
learning/ training loop.

**Learning &
Justifiable Trust**



**Good Outcomes:
Basis for Trust**



Sound multi-objective
decision-making

Explainability



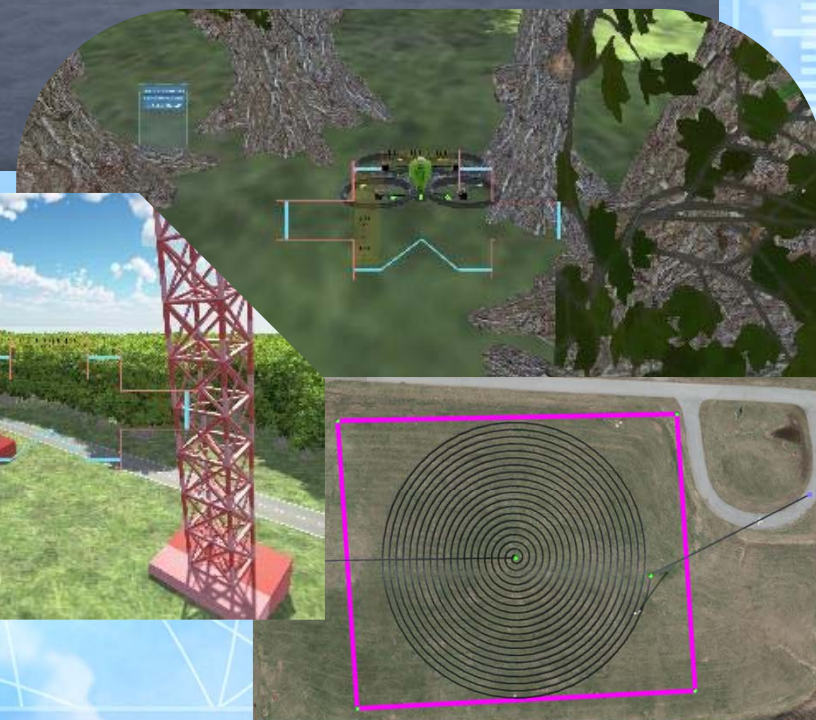
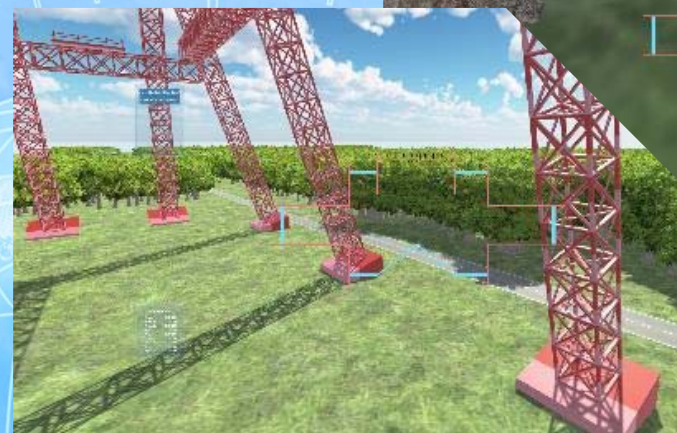
These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs.



Persistent “TwinSim” (AEON/BEAM)

Digital Twin Artifacts/Environments

“Digital Twin” Ecosystem

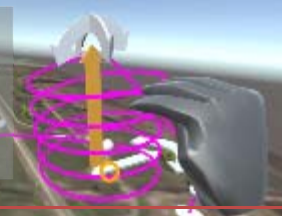




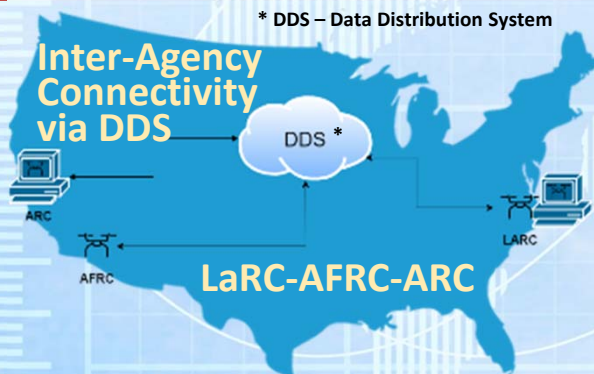
LaRC Autonomous Systems Simulator: AEON BEAM

Distributed Framework for Modeling, Simulation and Realization of Autonomous Systems

Natural & Immersive Interfaces for HAT



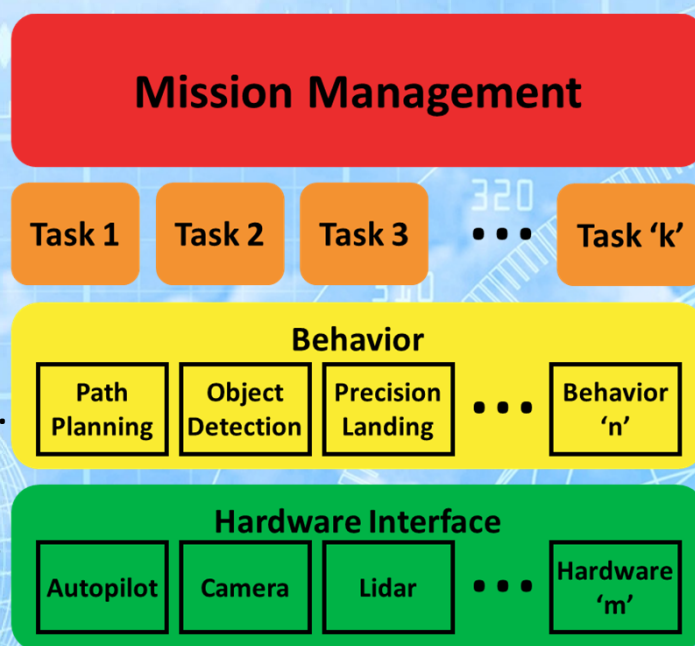
* DDS – Data Distribution System



Mixed Reality Sim-to-Flight

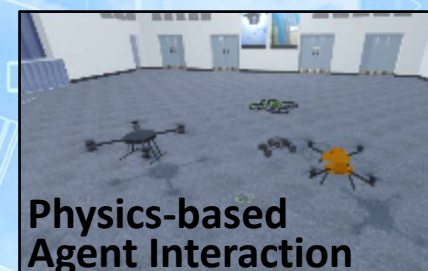
Autonomous Entity Operational Network (AEON)

- AEON is an architecture framework for sim-to-flight.
- AEON is a suite of libraries and self-contained applications supporting myriad features of critical autonomous capabilities.
- The AEON Framework is modular and easily extensible by developing new nodes to add system capabilities.



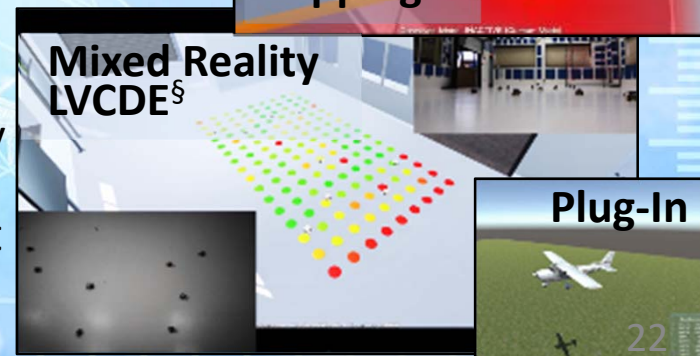
Baseline Environment for Autonomous Modeling (BEAM)

- Unity-based persistent sim that uses AEON and other frameworks
- Single- and Multi-Agent
- Mixed Reality for Data and Performance Analysis
- Multi-user with optional dependency on central servers. Also standalone.
- Modular plug-in capability
- Common communication interface (DDS) to support external research code



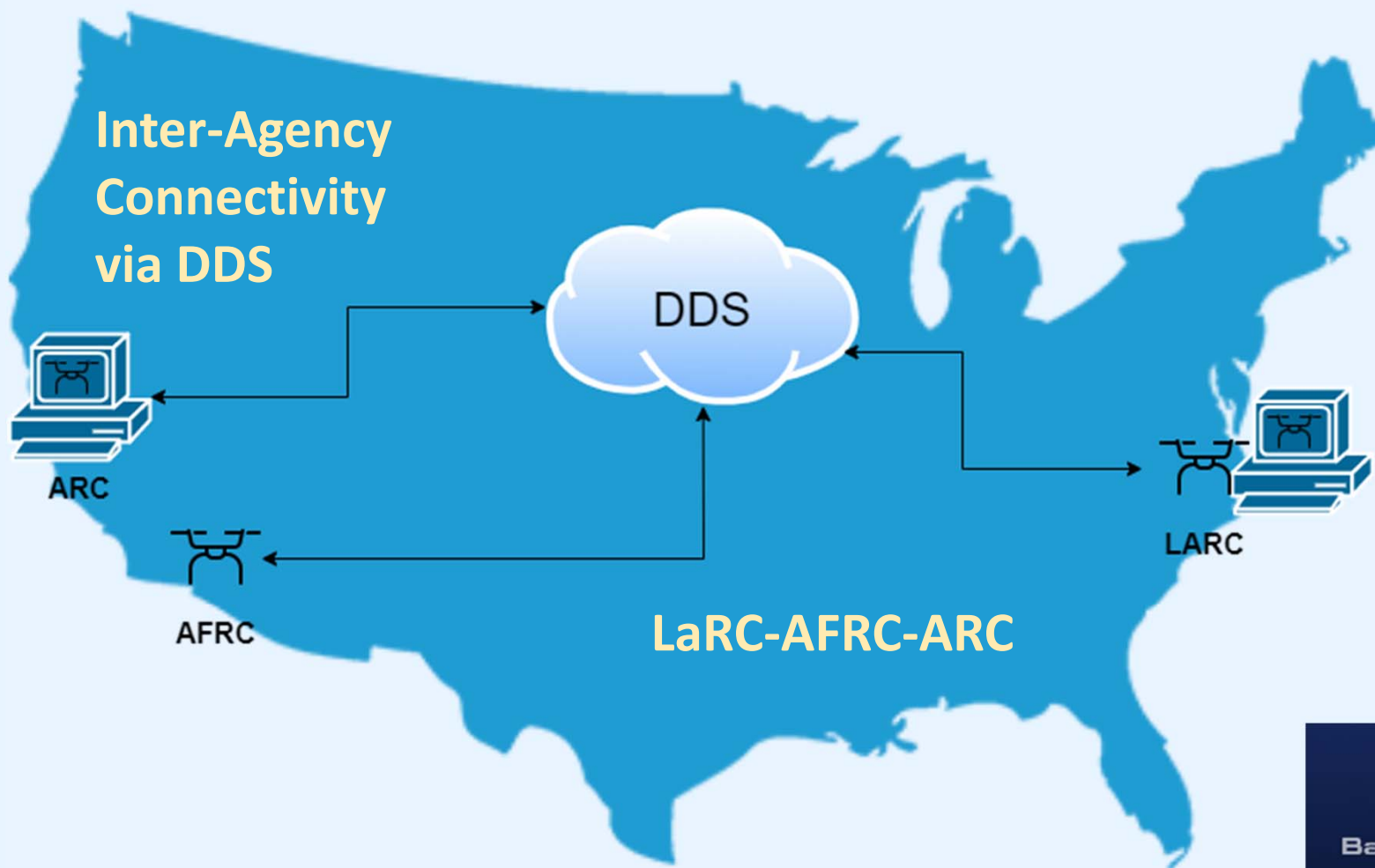
Lidar-based Mapping

Mixed Reality LVCDE[§]





BEAM Inter-Center Distributed Sim

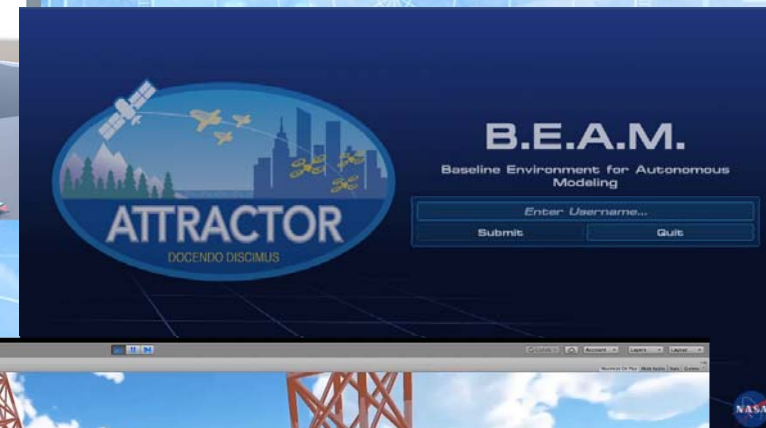
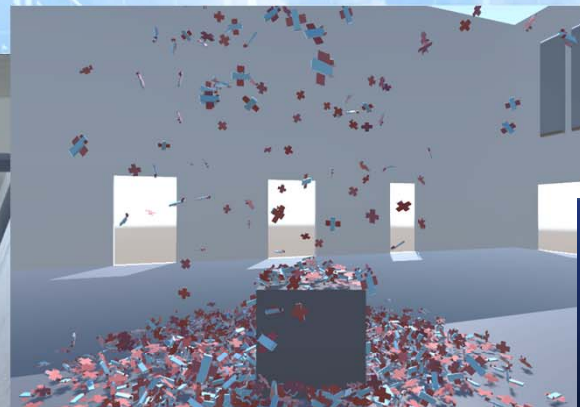


B.E.A.M.

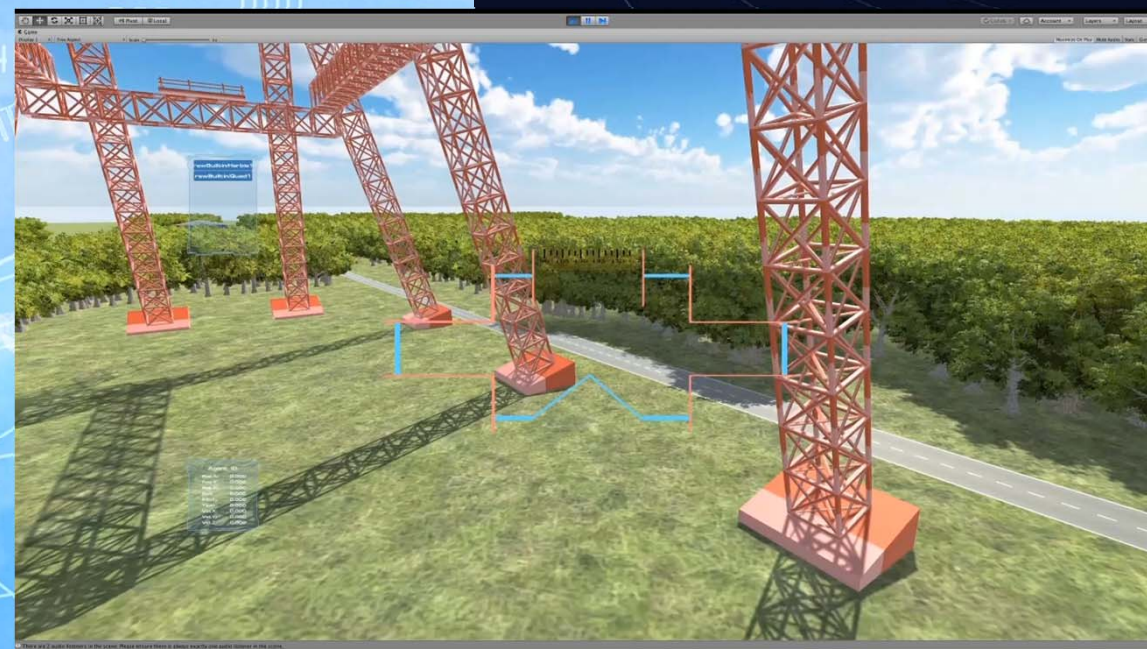
Baseline Environment for Autonomous Modeling



ATTRACTOR: BEAM

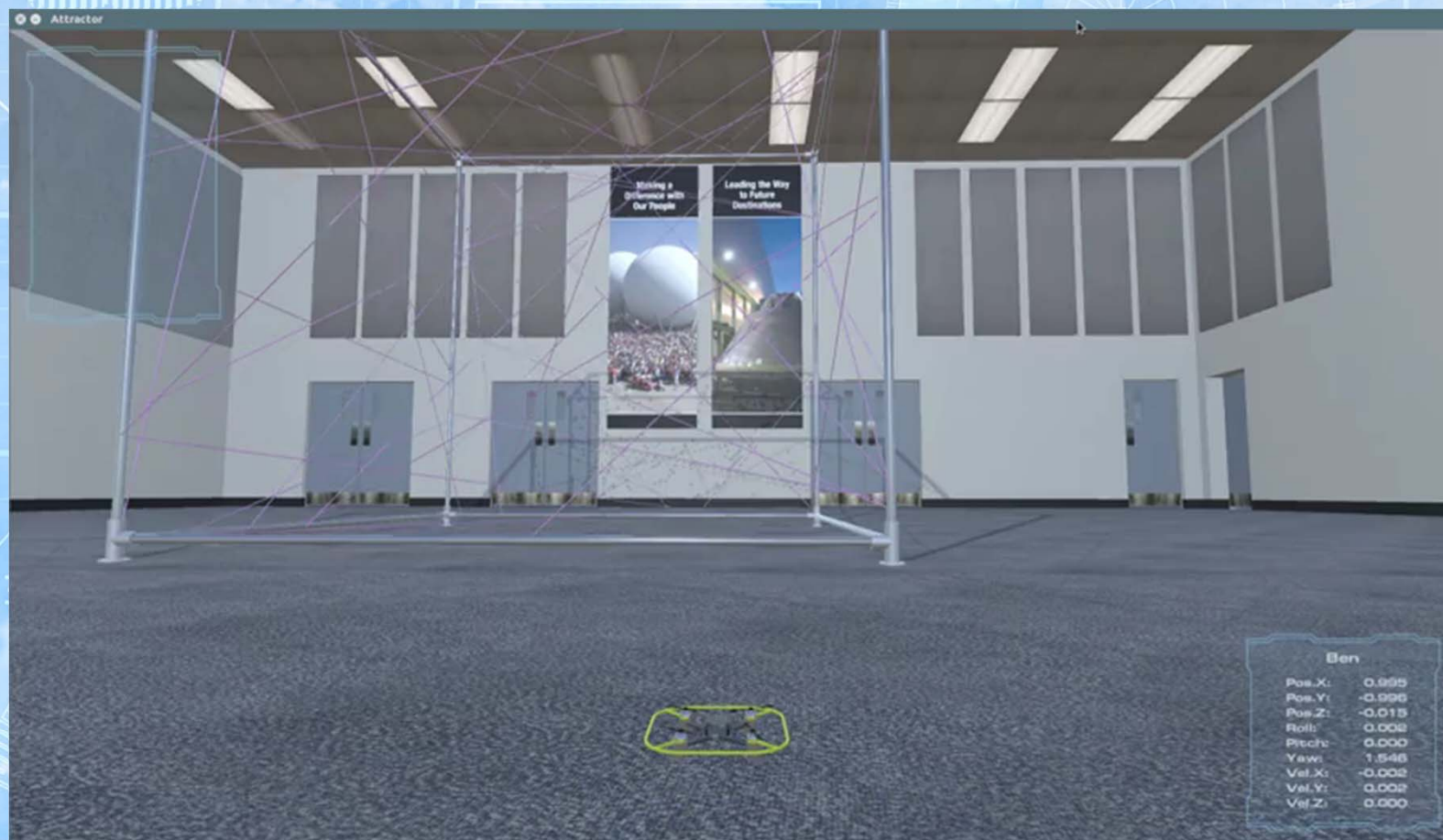
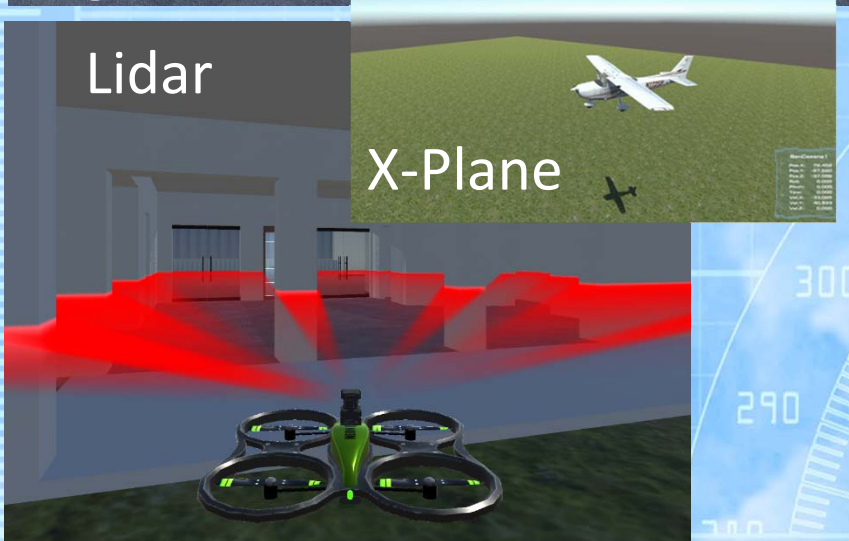


- Unity-based simulation environment
- Multi user
- No/Minimal dependency on central servers
- Modular plug-in capability
- Common communication interface (DDS) supporting external research code





BEAM Extensibility: Sensors, Environments, Vehicles, etc. — Ben Kelley



Maze environment with Crazyflie™ vehicle

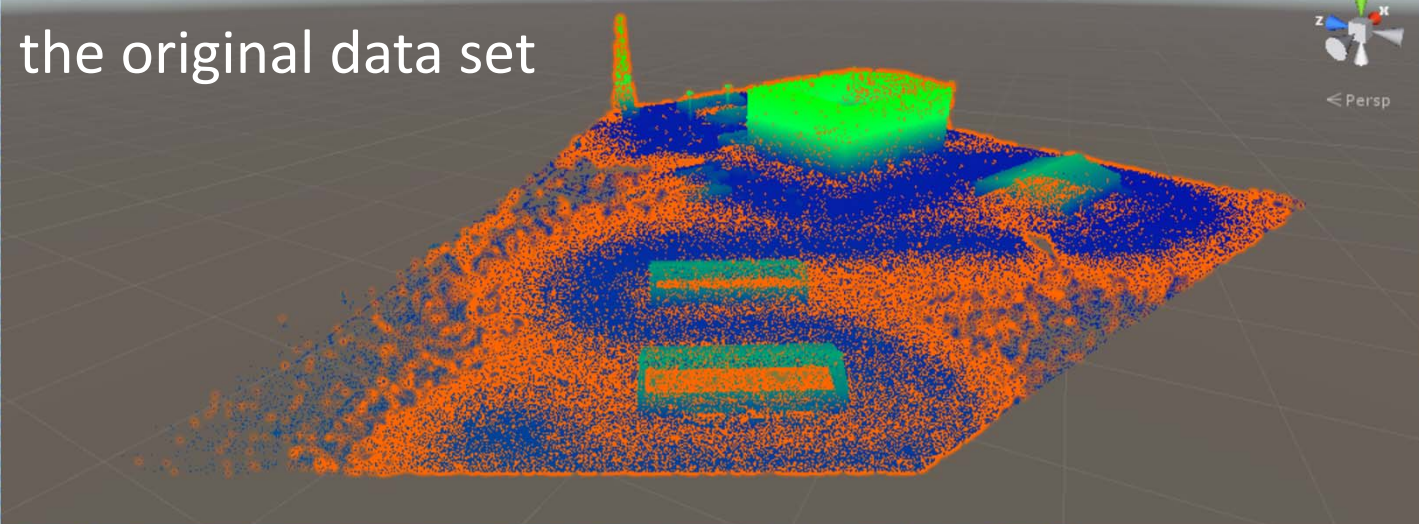




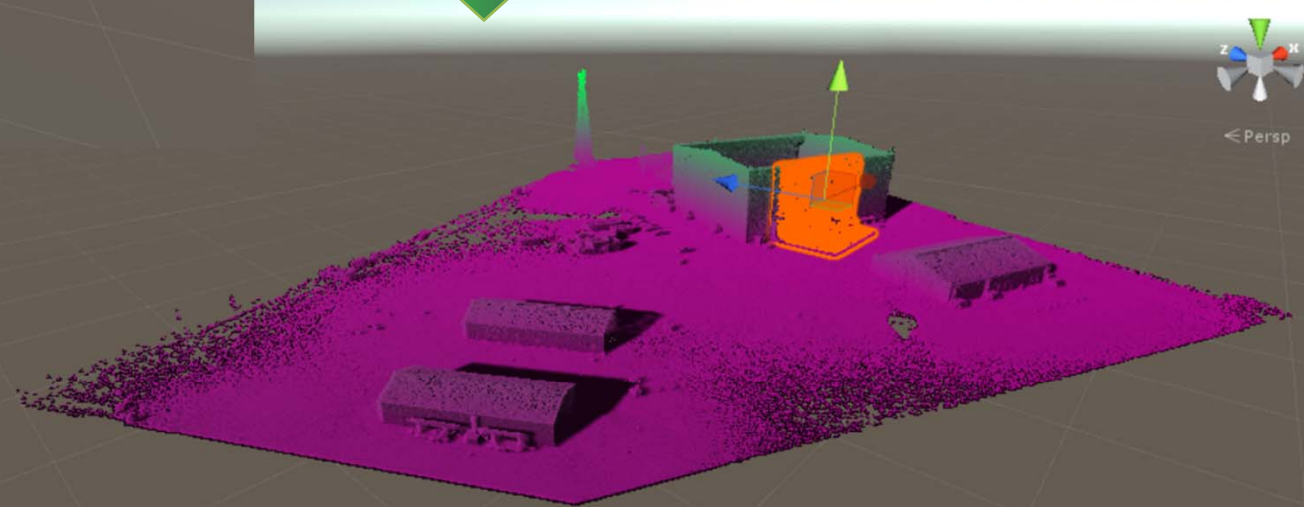
BEAM Extensibility: LIDAR Data

Tools: Data Tiles – Jason Holland

the original data set



Using a grid format,
the simulator can now load
nearest neighbor datasets
versus the entire data set for
an increase in performance



data organized in a grid format



Mixed-Reality for LVCDE





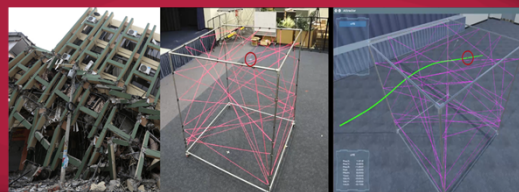
The ATTRACTOR Challenge:

Under The Hood: Trajectories

Multidisciplinary Component of
Complex Multi-agent Systems

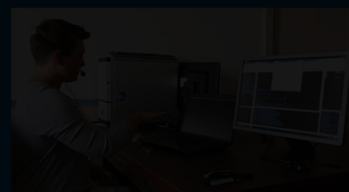
Measurable trustworthiness
of fundamental functional
components

Context:
Trajectories



Heterogeneous multi-agent (including human-machine)
teaming is essential

Teaming



Efficient two-way
learning/ training loop.

Learning &
Justifiable Trust



Good Outcomes:
Basis for Trust



Sound multi-objective
decision-making

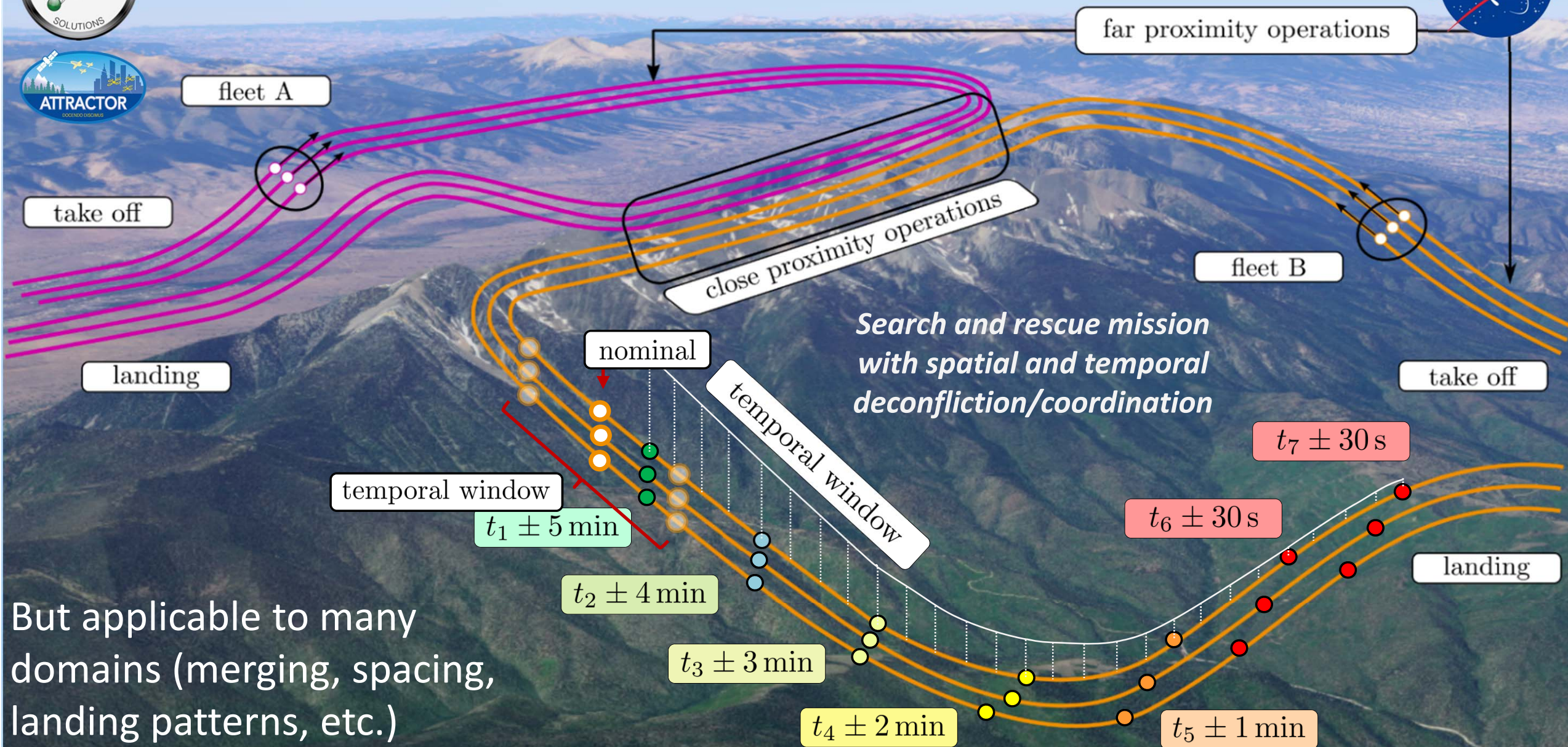
Explainability



These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs.

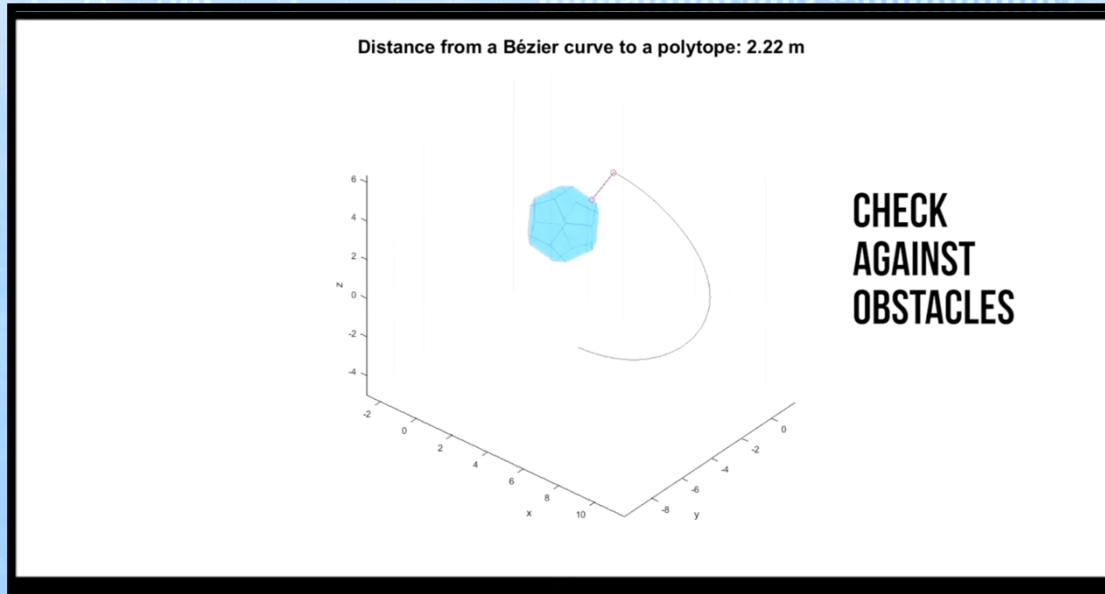


DRM: Search and Rescue operations in mountainous area





Bezier-Based Trajectories and Polytopes - Javier Puig-Navarro



Components

- Temporal Queries
- Proximity Queries
 - To limit the scope, obstacles are represented as polyhedrons
 - Trajectories are represented by a set of time parameterized polynomial (Bézier) curves

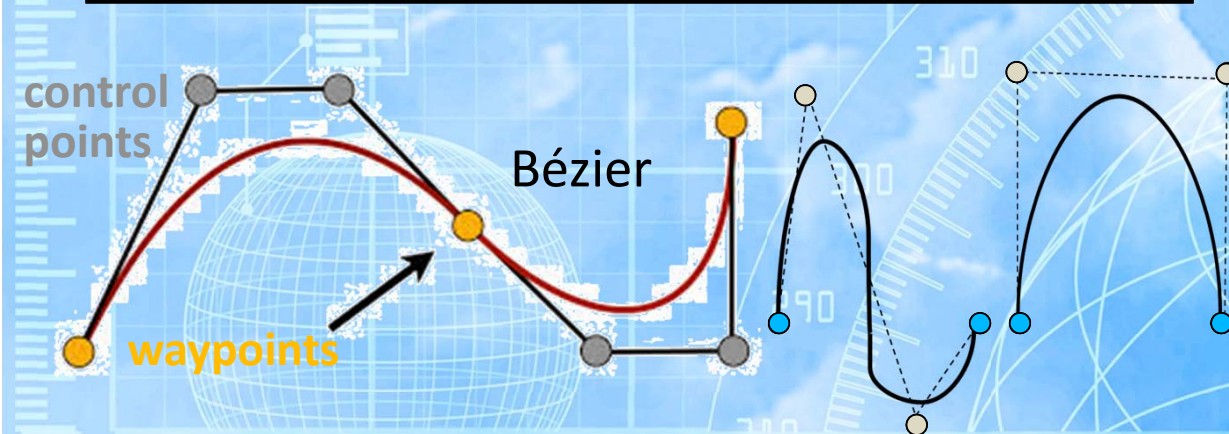
Enables

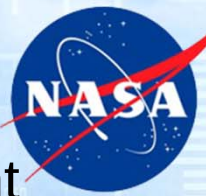
- Time Coordination/Deconfliction
- Spatial Coordination (Proximity Ops)
- Trajectory-Based Operations

Tolerance Verification = TQ + PQ

Convergence & Advancing the SOA:

- Tolerance Verification
- Silhouette Informed Planning



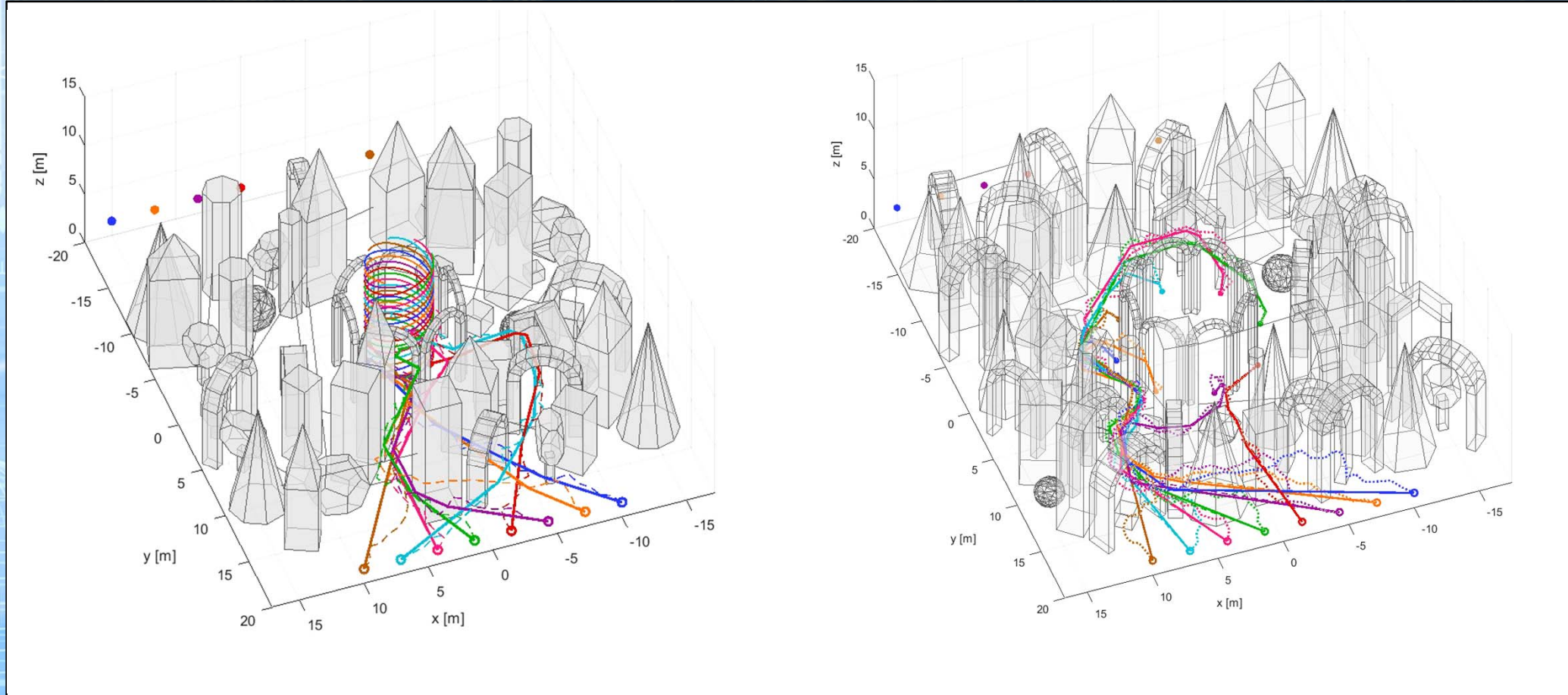


Advancing the SOA - Javier Puig-Navarro

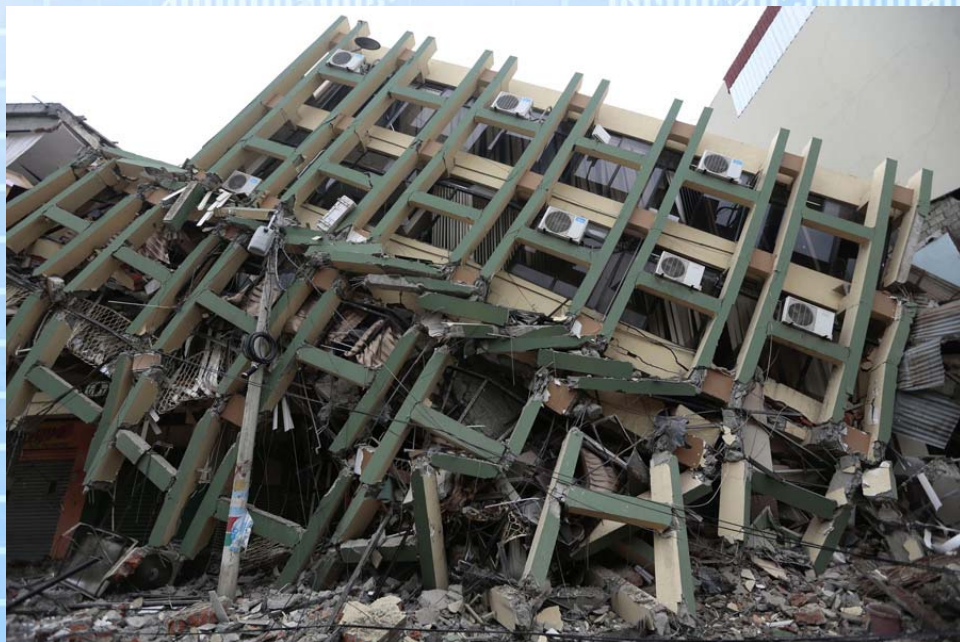
- *Adaptation of trajectory/path planning to local geometry of the environment* is a *unique* capability crucial in building a *trustworthy* autonomous system.
- The approach does not resort to curve sampling (e.g., discrete waypoints); it checks the entire curve. As a result, it is *scalable* – a *requirement for multi-vehicle operations*. It is **20 times faster** than distance computation, on average.

TBO

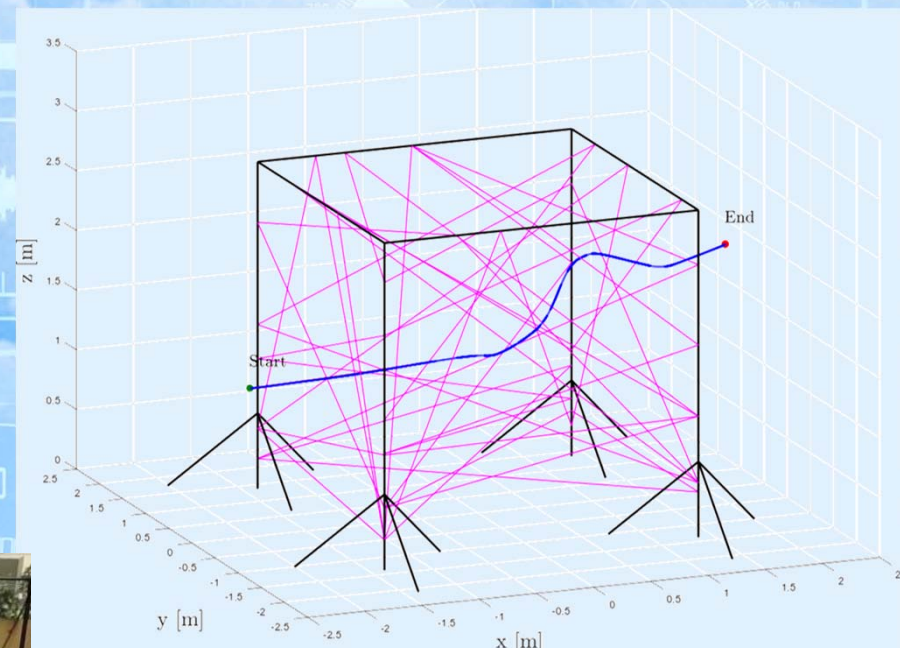
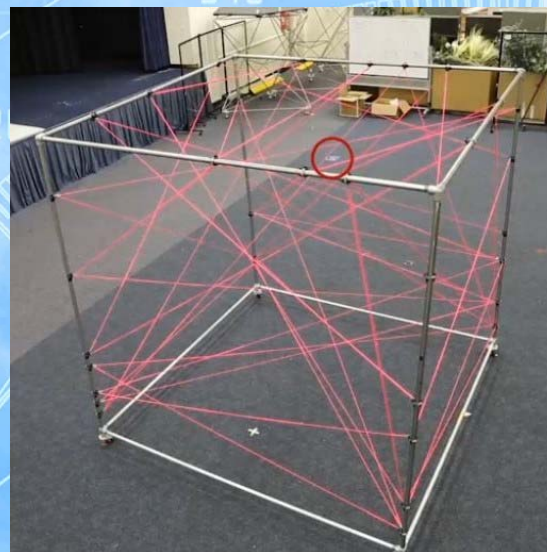
Silhouette-Informed
Tolerance Verification



The Wire Maze



Effects of an earthquake in Ecuador, April 2016. Image credit: Juan Cevallos/AFP/ Getty

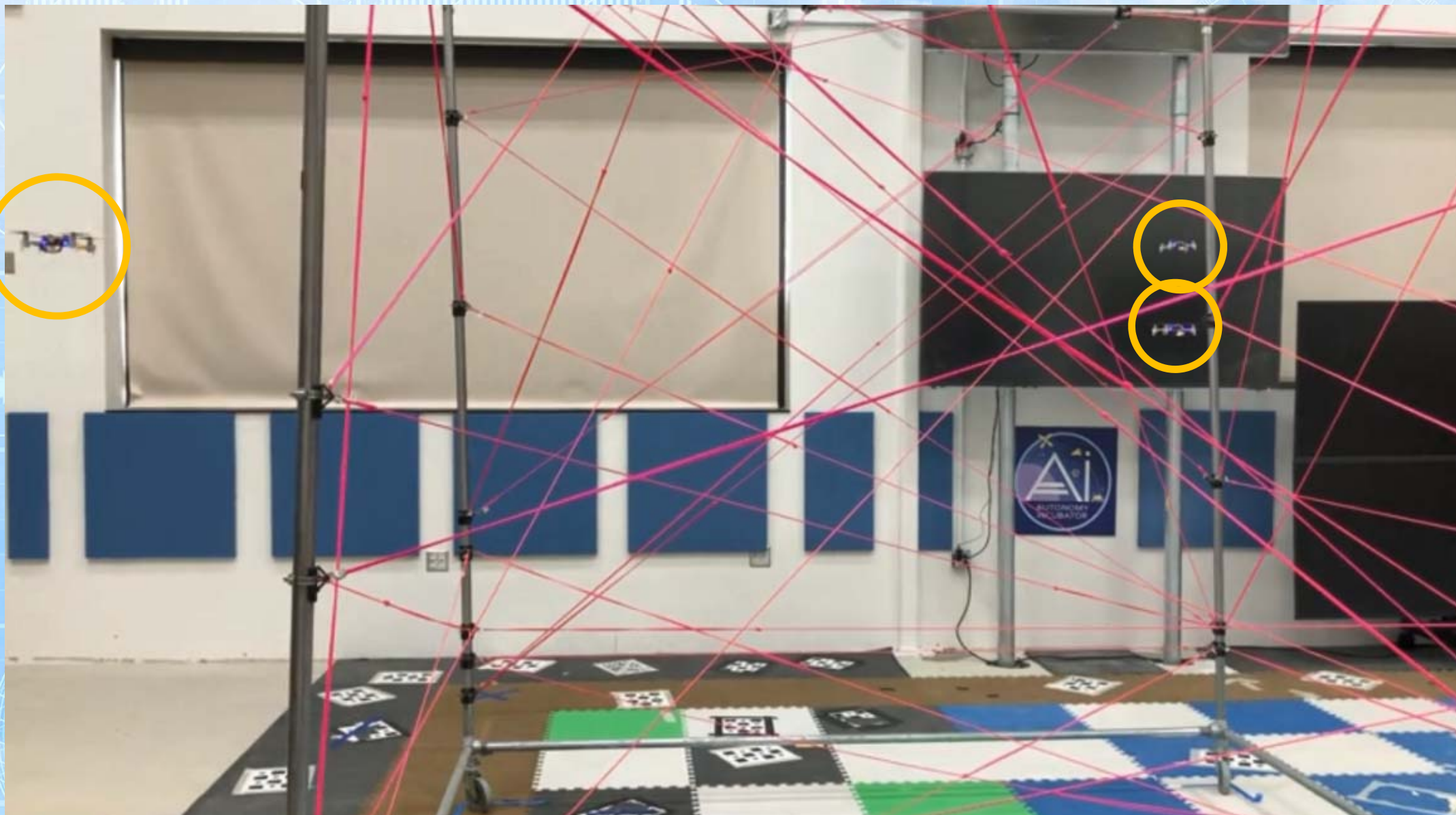


An analog for a cluttered environment in simulation...

and in reality.



Wire Maze Flight Demo



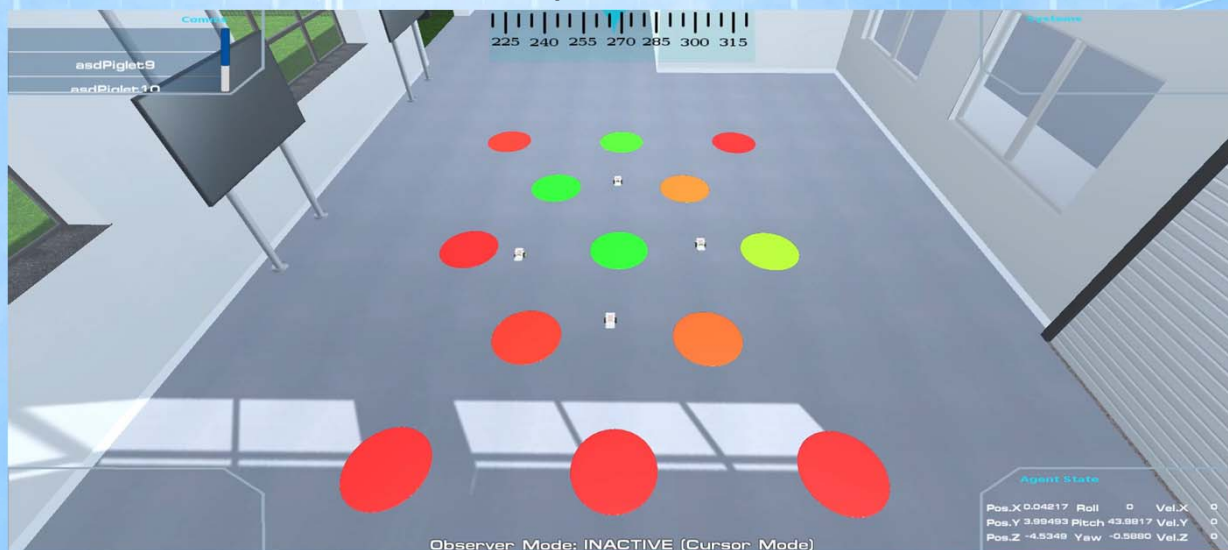


Multi-Agent Guidance: Potential Fields

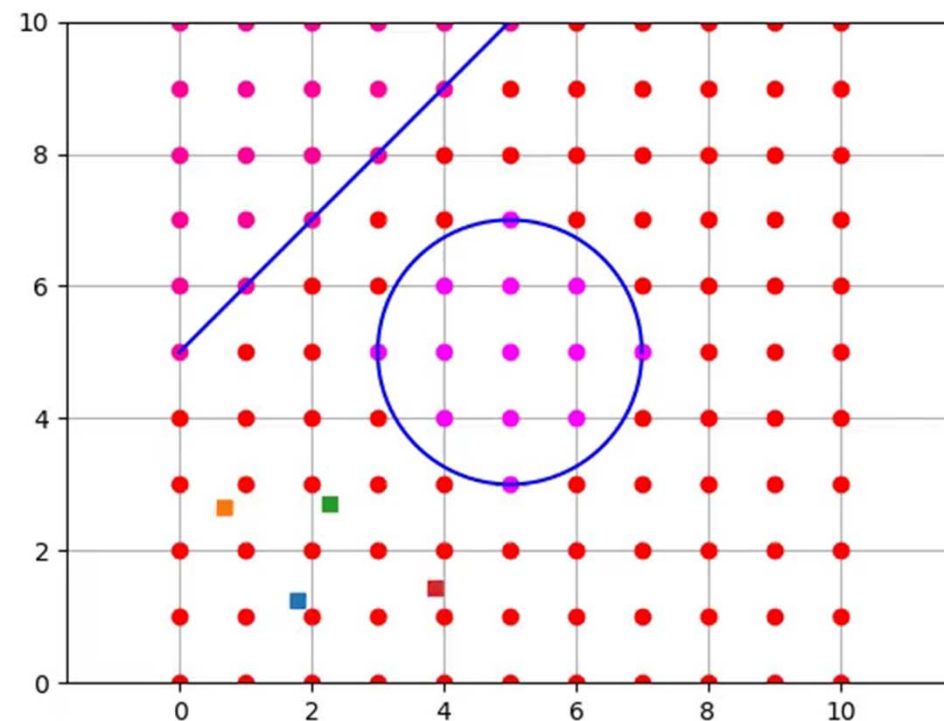
John Cooper



- New application of “old” ATM methods
- Generate velocity commands using gradient of potential field
- Points represent centers of search regions
 - Gaussian radial basis functions around each point
 - Vertices can be evenly distributed as a grid or have variable density



**Hardware-in-the loop (HITL) demonstration:
Integration with persistent simulation**



**Clutter Thresholding: agents recognize
regions that are not penetrable**

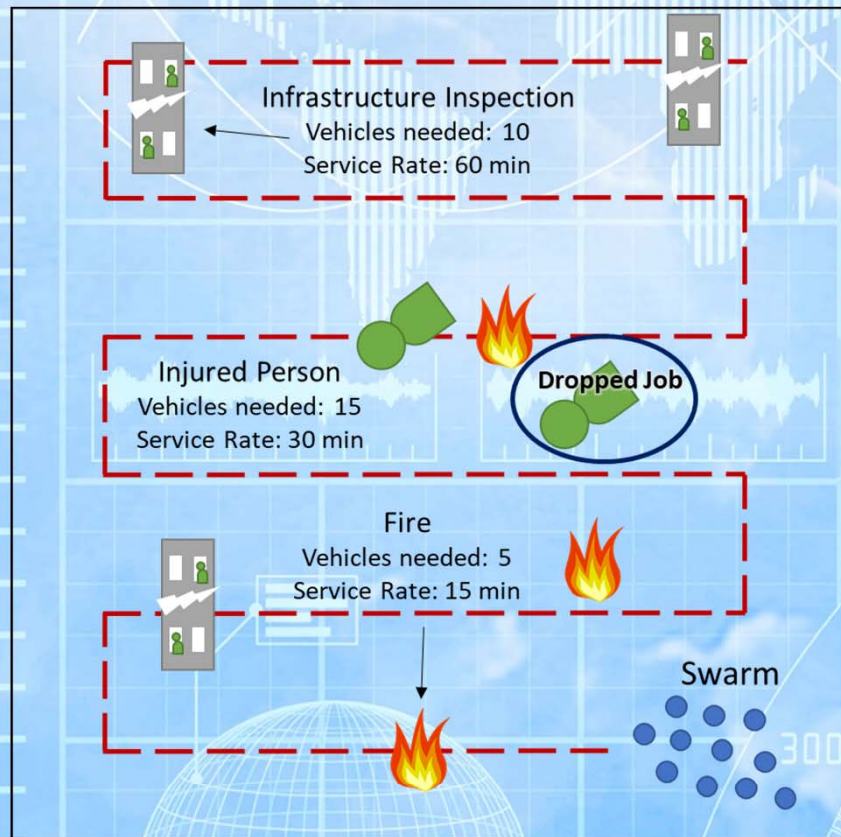
John Cooper & Danette Allen



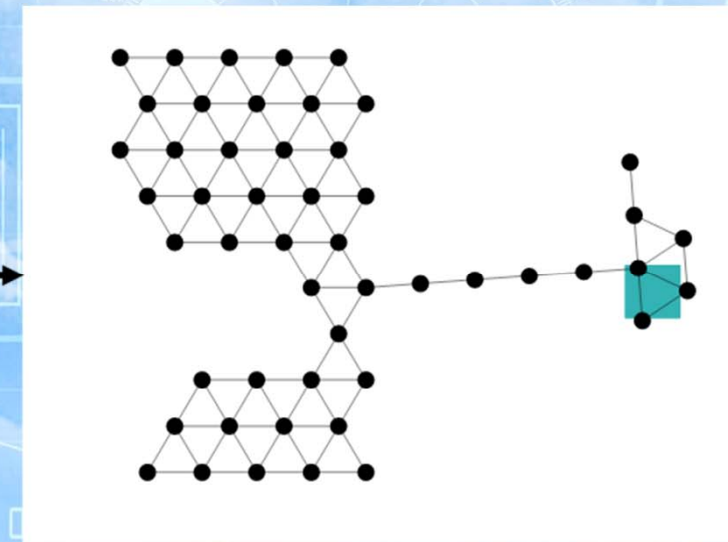
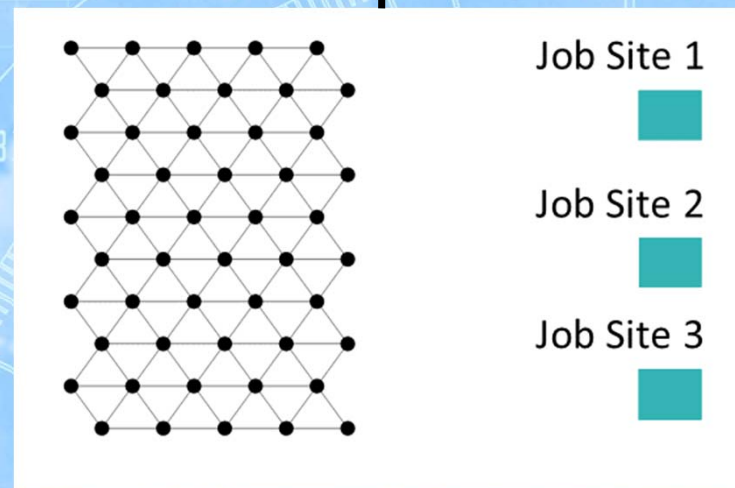
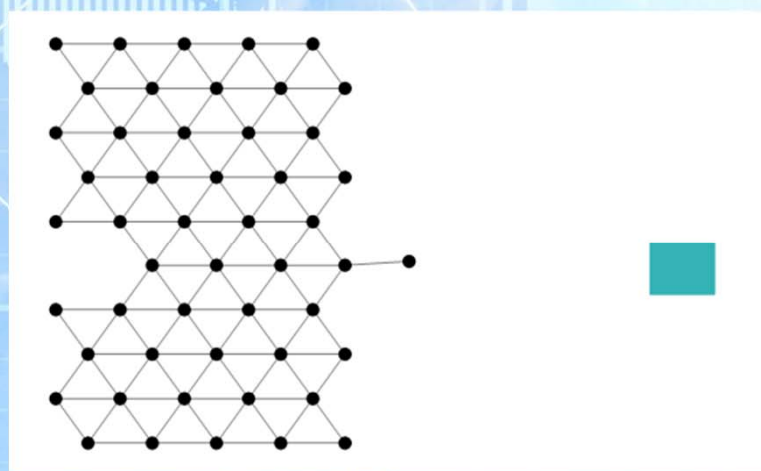
Swarm Mission Planning and Execution

Meghan Chandarana

Swarms: a multi-agent system



Swarm Search & Service (SSS) Missions



Goal: Send robots to each site to accomplish job.

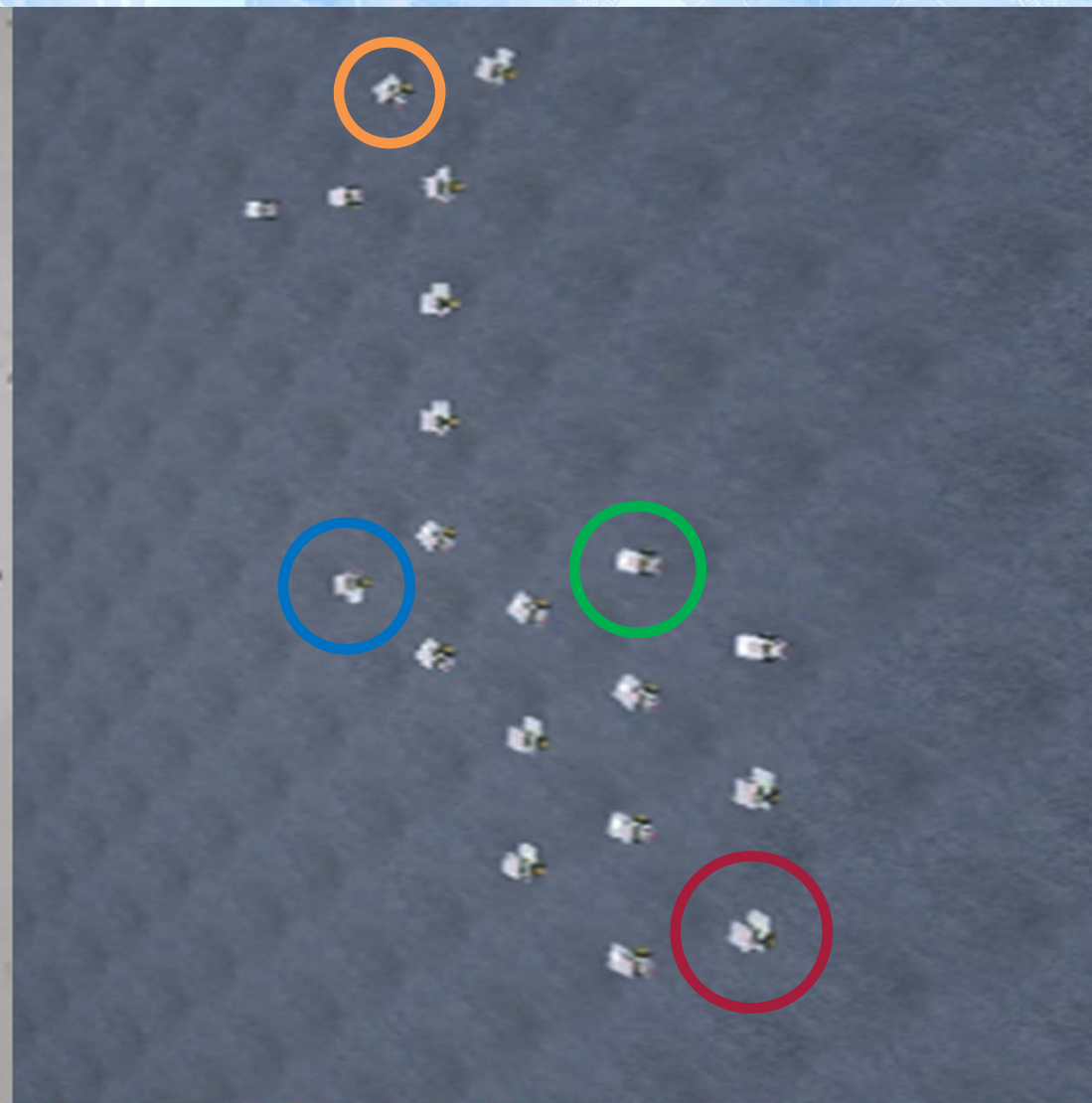
Method: Push a single line of robots out of the swarm

- Maintain connectivity with main swarm
- Allows for multi job sites to be simultaneously serviced



Demo: Sub-swarm Breakoff and Rejoin

Meghan Chandarana



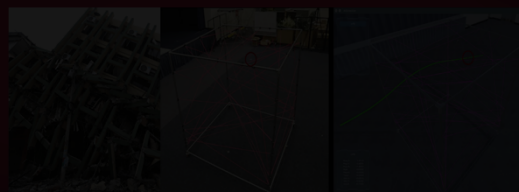


The ATTRACTOR Challenge: Human-Machine Teaming

Multidisciplinary Component of
Complex Multi-agent Systems

Measurable trustworthiness
of fundamental functional
components

Context:
Trajectories
and DRM



Sound multi-objective
decision-making

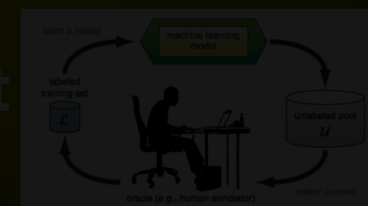
Heterogeneous multi-agent (including human-machine)
teaming is essential



Teaming

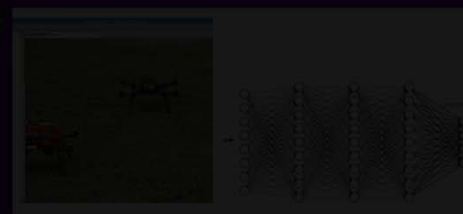
Efficient two-way
learning/ training loop.

Learning &
Justifiable Trust



Good Outcomes:
Basis for Trust

Explainability



These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs.



Human Machine Teaming



HOLII GRAILLE

Idea: Engender trust in human-machine teaming via natural human-machine interaction

- speech
- gestures,
- mixed-reality
- etc.

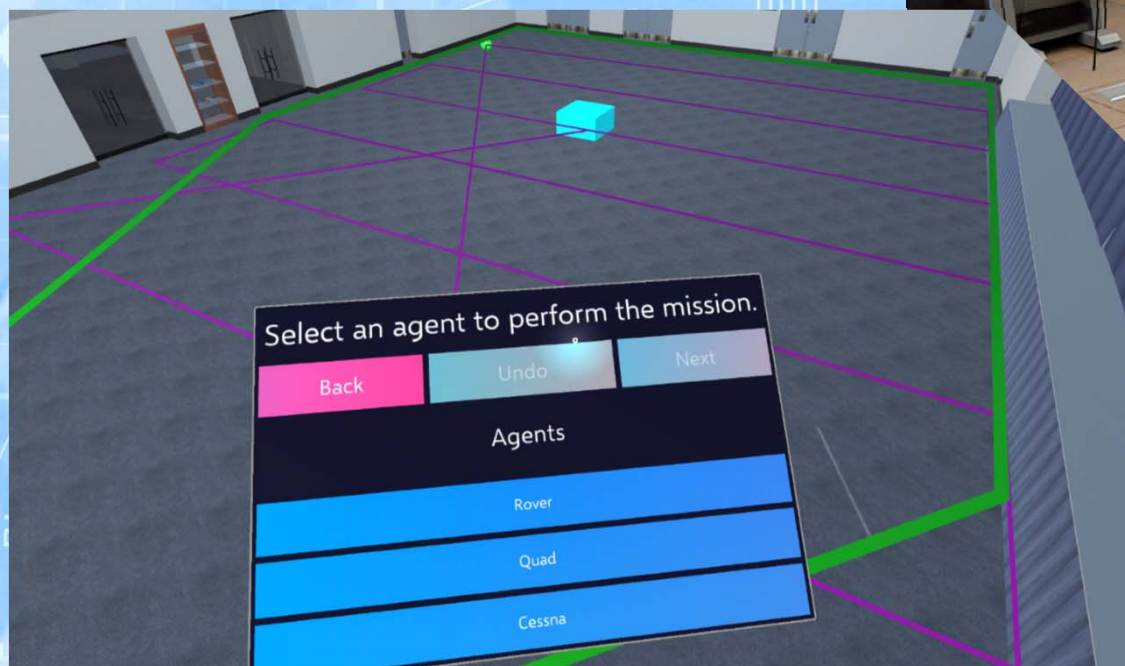
“Put that there”





Mixed Reality with HoloLens

Feasibility Study - Erik Coltey

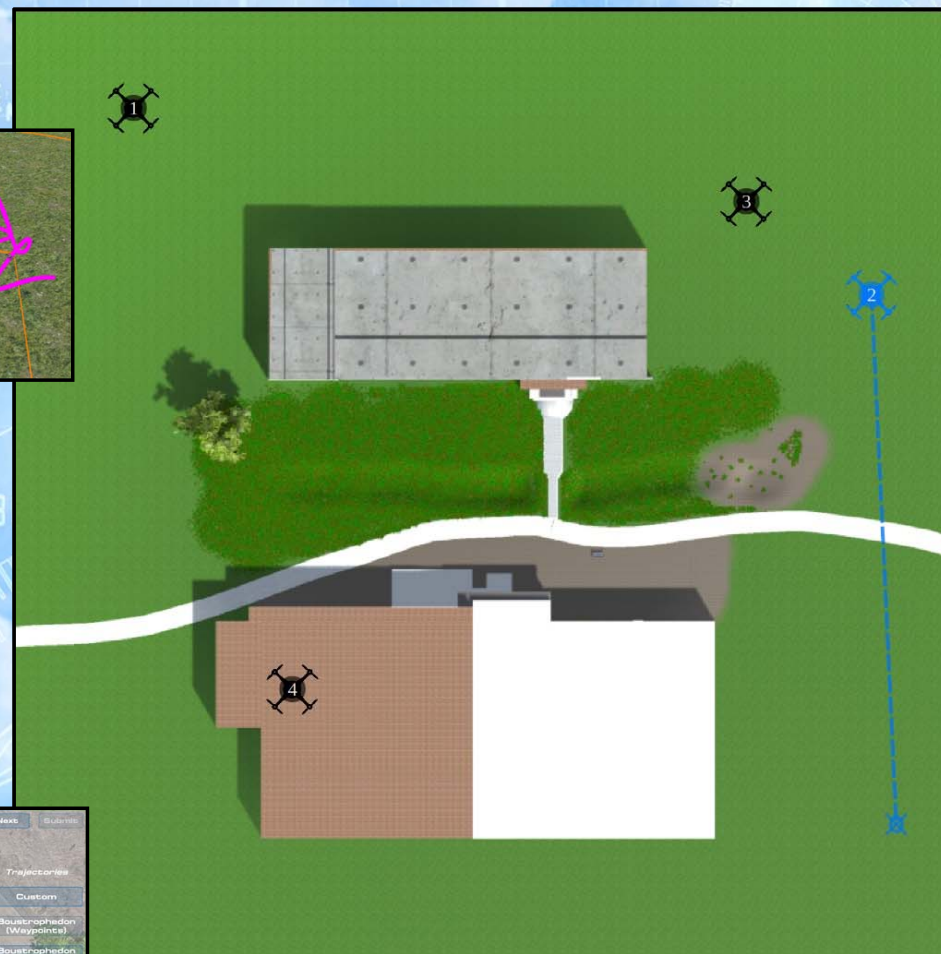




AMELIA - Bryan Barrows

Autonomous Mission Experimental Logistics Interactive Assistant

- Multi-modal mission planning assisted by intelligent systems
- Preference of input modality
 - Natural
 - Performance
 - Trust
- Effective or invasive autonomy?
- Feedback methods for users to improve experience



Mission Complete

Time Remaining: 1:53

The 'Amelia' speech indicator was not detected.



Amelia Helper

Amelia: Some assets are positioned too closely. Would you like to move this asset?

Yes

No



Individual Differences

Lisa Le Vie

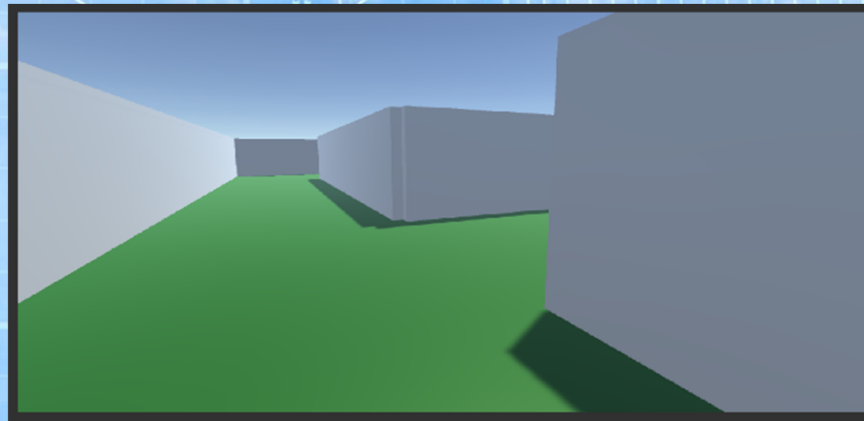
- What should agents say (not say) and when? And how?
- Maze Experiment: “ ‘When’ and ‘how often’ do people ask for help from AI ”

Level of detail

“The first right or either of the left turns will lead you to many dead ends, the second right will eventually lead you to the goal”

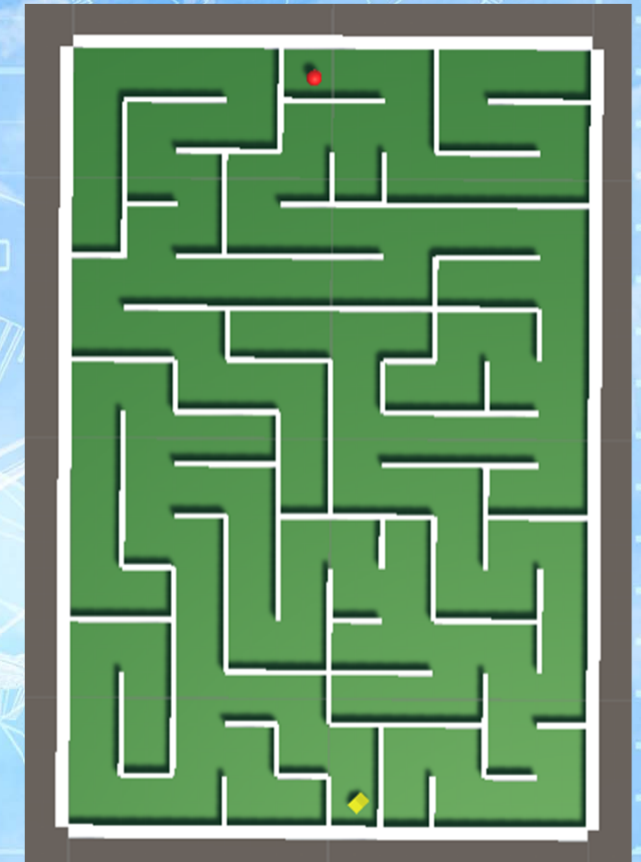
or

“Second right”



Summer 2019 Study –

- 13 participants
- 38% never asked for help (2:29)
- 1st request at 3:11 (5:53)
- Varied level of detail
- Need more data



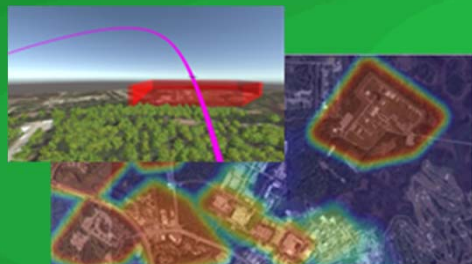


Toward Trustworthiness & Trust (T&T) in Autonomous CPH Systems

Multidisciplinary Components of
Complex Multi-agent Systems

Measurable trustworthiness
of fundamental functional
components

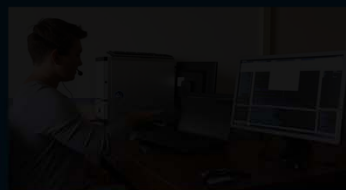
Context:
Trajectories
and DRM



Sound multi-objective
decision-making

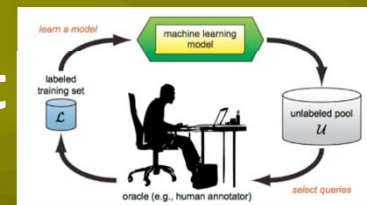
Heterogeneous multi-agent (including human-machine)
teaming is essential

Teaming



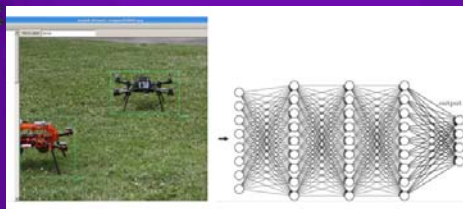
Efficient two-way
learning/ training loop.

Learning &
Justifiable Trust



Good Outcomes:
Basis for Trust

Explainability



These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs.



T&T

- T&T: attributes of system and participants
- The remainder of the presentation describes the components of ATTRACTOR that support decision-making in training and operation of the system.

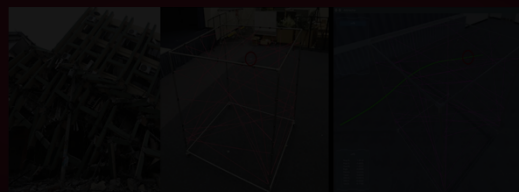


Machine Learning (ML): Explainable AI ML-enabled Safety-critical Systems

Multidisciplinary Components of
Complex Multi-agent Systems

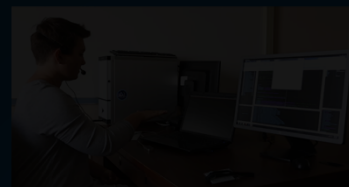
Measurable trustworthiness
of fundamental functional
components

Context:
Trajectories
and DRM



Sound multi-objective
decision-making

Heterogeneous multi-agent (including human-machine)
teaming is essential



Teaming

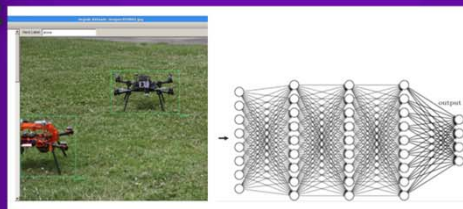
Efficient two-way
learning/ training loop.



Learning &
Justifiable Trust

Good Outcomes:
Basis for Trust

Explainability



These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs.

Motivation

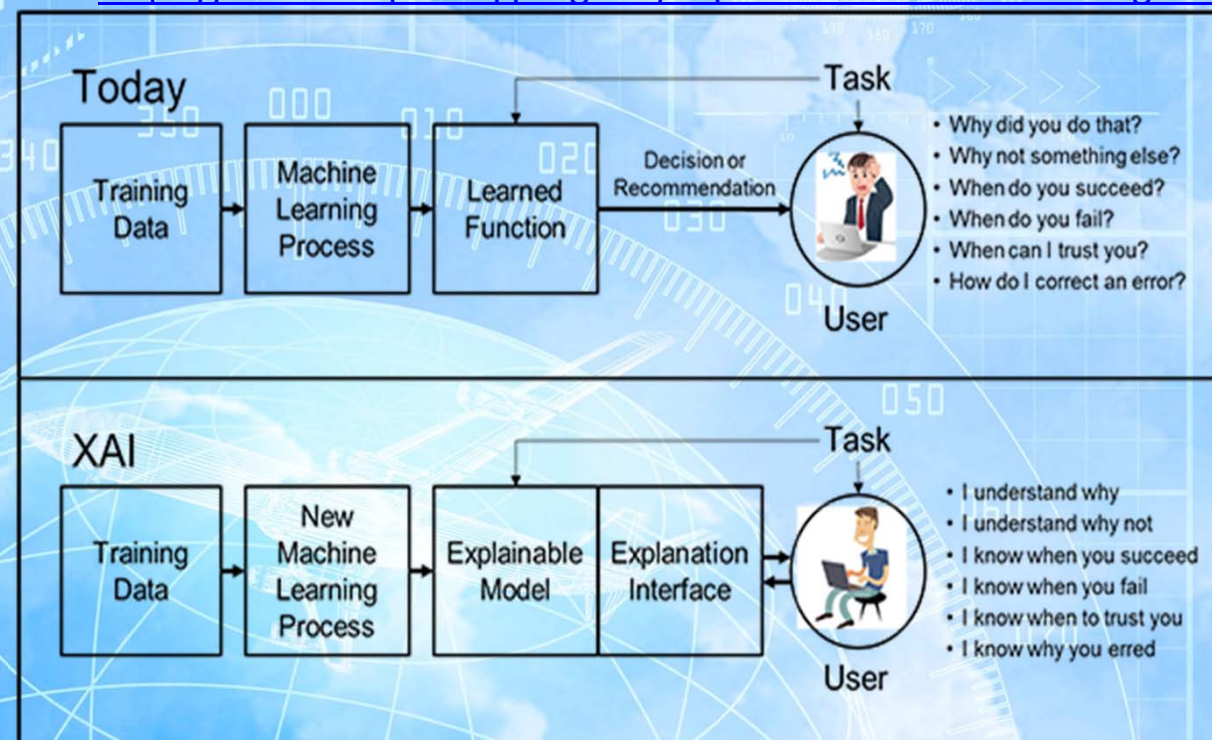
- ML is fragile. Why ML in safety-critical environments?
 - Situational awareness with sensors that require interpretation (perception) relies on ML
 - ML can also serve in an advisory capacity

<https://www.darpa.mil/program/explainable-artificial-intelligence>

- Why XAI?

- Understand decision-making processes; debug; train; a posteriori analysis
- Humans prefer decisions with explanations in accessible terms

↑ T&T

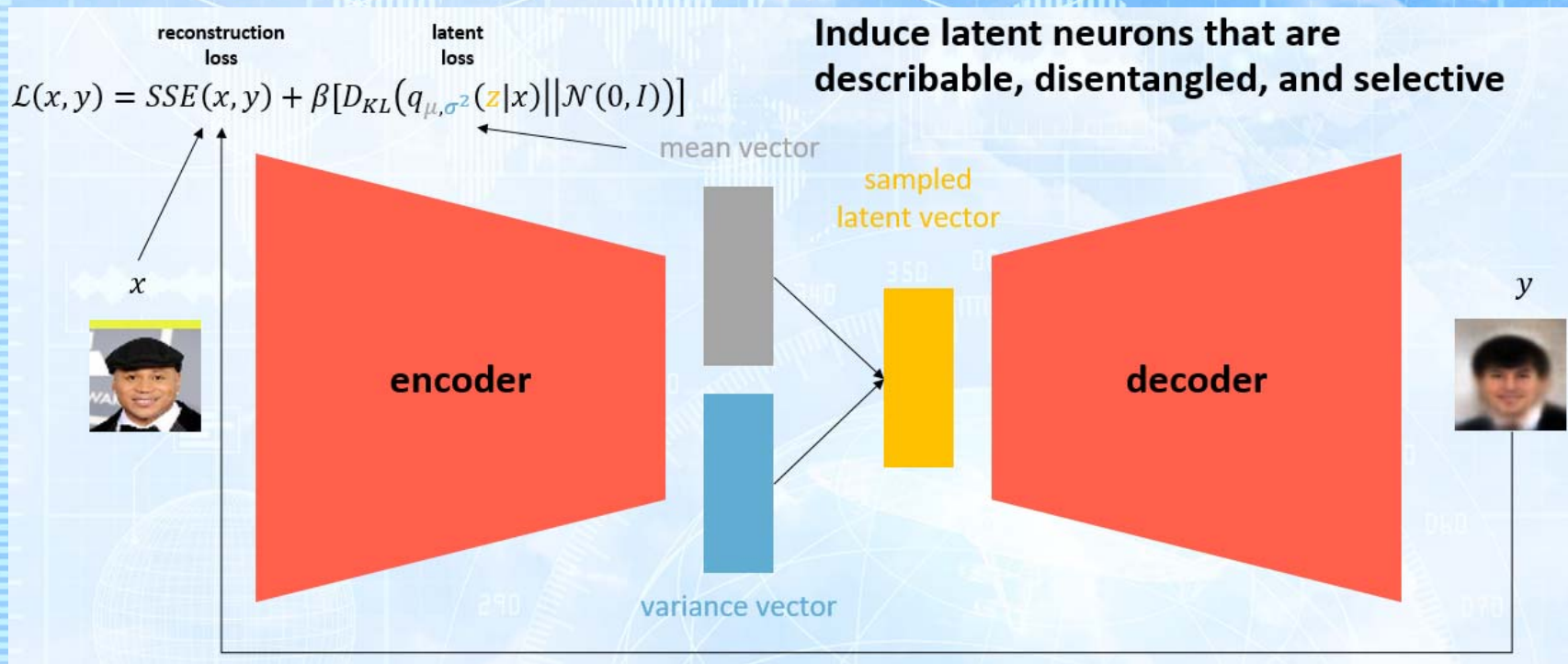


Explaining ML with Variational Autoencoders (VAE)

Loc Tran

- Idea/Concept: Is there a network neuron with high selectivity to a feature that can be clearly described (e.g., is there a 'red shirt' neuron?) and can we overcome entanglement?

The β -Variational Autoencoder places weight on disentanglement



- Autoencoders are neural networks that learn representation of data
- Autoencoders compress and reconstruct input images
- β -VAE goal: learn disentangled latent parameters from unsupervised data
- MOISTURIZE Tool: developed in house to explore implementation of VAE

Search and Rescue (SAR) Mission

Deploy UAV



Search Phase

Search Path



Detect Person



Move in closer



Detect Face

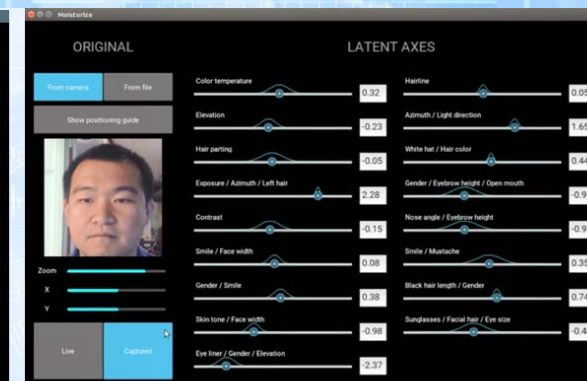
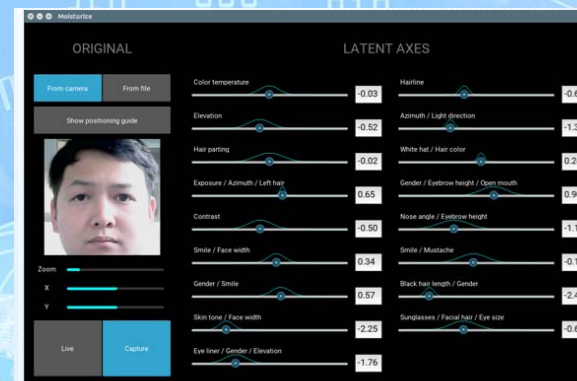
VAE

latent representation

2.4
-1.8
7.2
8.7
-9.1
5.5

Findings:

- VAE are a feasible tool for representing human interpretable features from complex data sets
- VAE do not produce completely disentangled representations
- Suitable as additional info or advisor; not for direct safety-critical action



Missing Person

VAE

latent representation

2.4
-1.8
7.2
0.7
-9.1
5.5



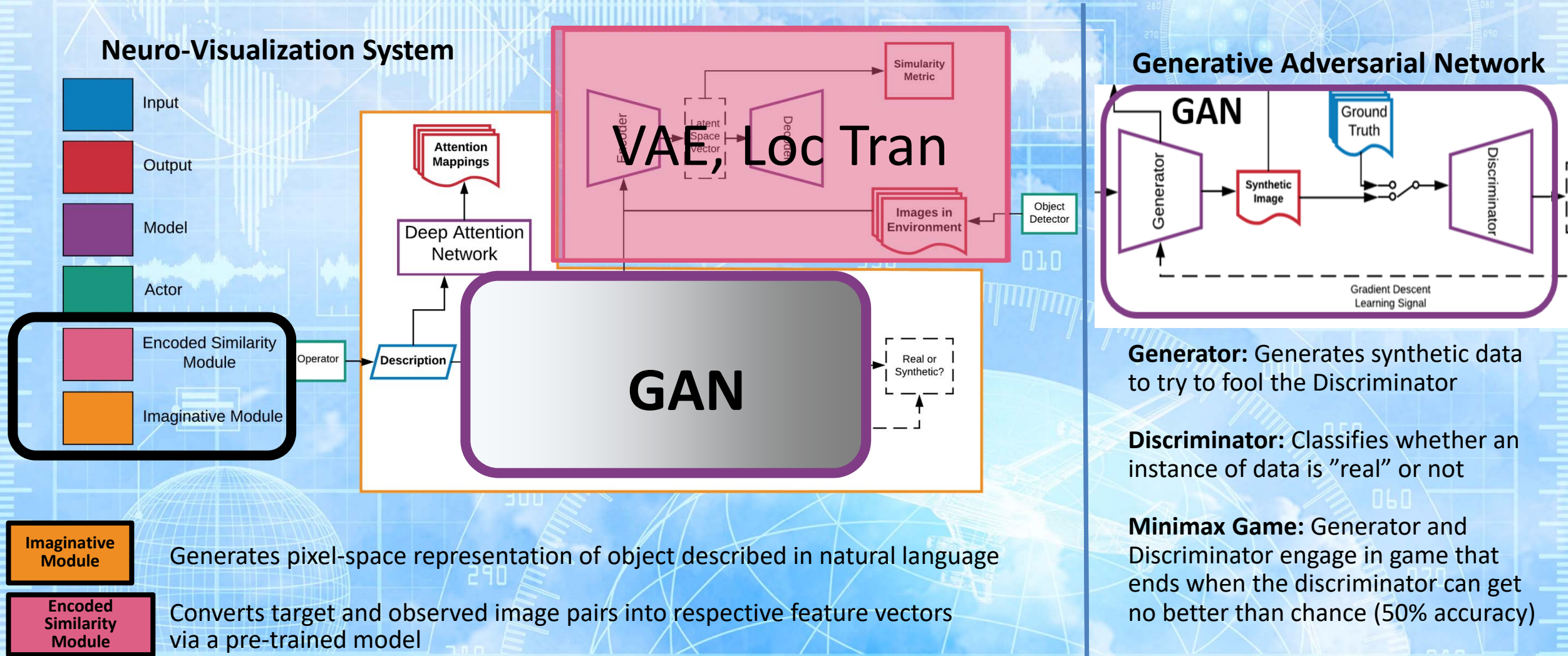
Detected Face

VAE

latent representation

2.4
-1.8
7.2
8.7
-9.1
5.5

- Idea/Concept: Describe an object of interest via neuro-visualization: tell a machine what it is looking for



Neuro-visualization: Examples of Findings

Birds



Good results rely on description context and localization of object in the training data

Humans



COCO dataset: Descriptions contain information holistic to the image, with low object specificity

Machine Imagination



Synthetic

Ground Truth

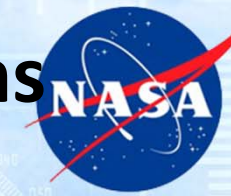
this is a small black bird with a white spot on its nape, it has a very large head and bill for its body.

Reconstruction from ground truth shows highly accurate but imperfect recall

Findings:

- Given properly configured data, this system can reliably generate explainable images of the subject classes
- Generated images, if not accurate as a whole, possess a collection of features meaningful to the algorithm
- Shows promise for operational xAI in the context of SAR; feasible as advisor to decision-maker





Improving Trust in Machine Perception through Explanations

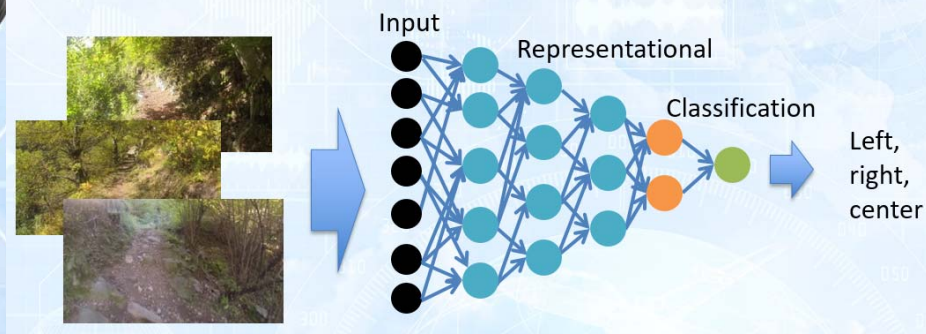
Adrian Agogino, Ritchie Lee, Dimitra Giannakopoulou

- Idea/Concept: Use explanation by example, nearest neighbors (KNN), in latent space for transparent and understandable decision making
- KNN is a viable but more intuitive alternative to standard decision-making for deep neural networks
- Test Domain: Neural Networks for Forest Trail Image Classification
 - UAS navigates with front facing camera; determines direction to follow trail

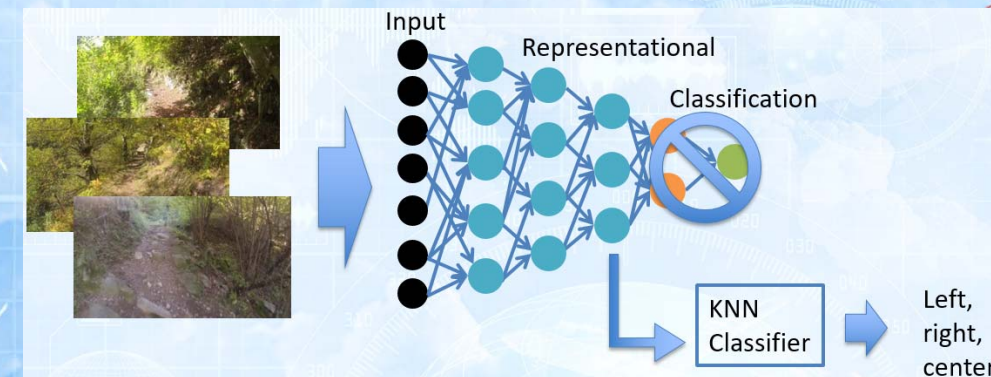


KNN, cont.

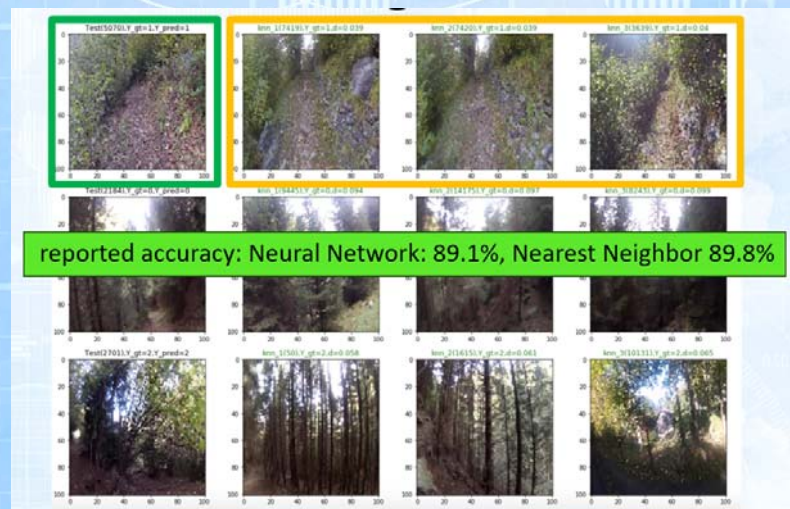
Deep NN Classification



K-Nearest Neighbor

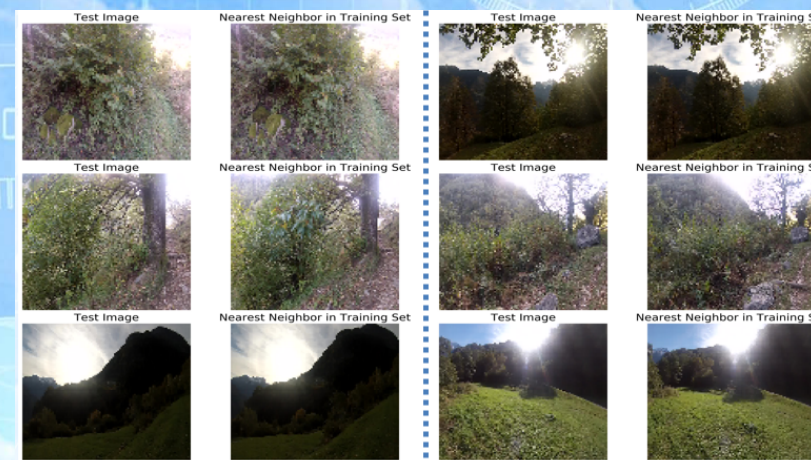


3-Nearest Neighbor Classifier (cosine similarity)



KNN identified unknown problem in time series data
Corrections were made to give more realistic performance measure

Time-Series Images



Findings:

- Performance similar to SOA deep NN
- Nearest neighbors help detect errors in training
- Visualization of neighbors indicates quality of distance functions
- Decisions are transparent
- Explanations are intuitive since they relate decisions to training data



Trustworthiness of ML-enabled Safety-critical Systems



Alwyn Goodloe

Findings:

- At current SOA, it may not be possible to write specifications of the function correctness for SOA ML systems. This makes them not amenable to traditional software verification techniques.
- Basic research remains to be done on how to verify functional correctness of ML enabled systems.
- A number of researchers are investigating verifying correctness properties of neural networks.
- For the foreseeable future, runtime verification will have to be used extensively. However, runtime verification faces the same challenge as formal verification approaches: the need to specify the property being verified.
- Perception poses a special problem because properties amenable to formalization are highly specialized and have yet to be invented for perception.

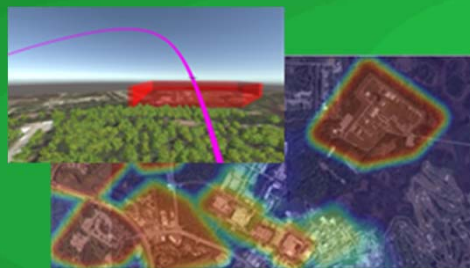


Supporting T&T System Design and Operations

Multidisciplinary Components of
Complex Multi-agent Systems

Measurable trustworthiness
of fundamental functional
components

Context:
Trajectories
and DRM



Sound multi-objective
decision-making

Good Outcomes:
Basis for Trust

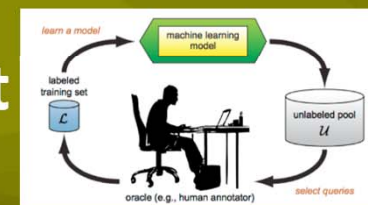
Heterogeneous multi-agent (including human-machine)
teaming is essential



Teaming

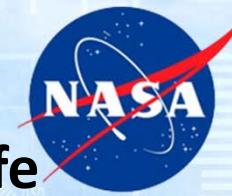
Efficient two-way
learning/ training loop.

Learning &
Justifiable Trust



Explainability

These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs

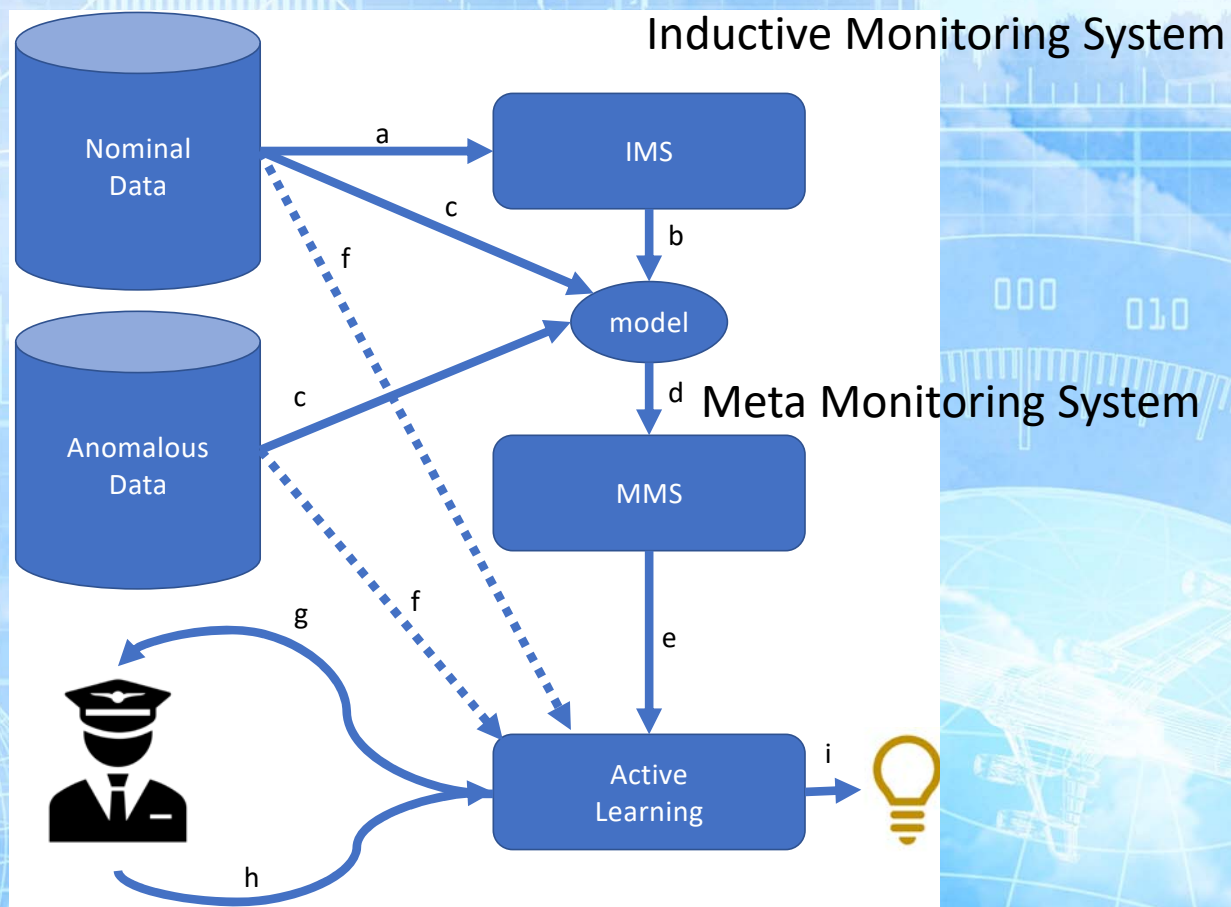


Incorporating Human Knowledge in AS through ML

Nikunj Oza, Kevin Bradner, David Iverson, Adwait Sahasrabhojane, Shawn Wolfe

- Idea/Concept: Facilitate trust in autonomous systems by using domain expert knowledge to identify anomalies and their precursors. Explain anomalies during operations.

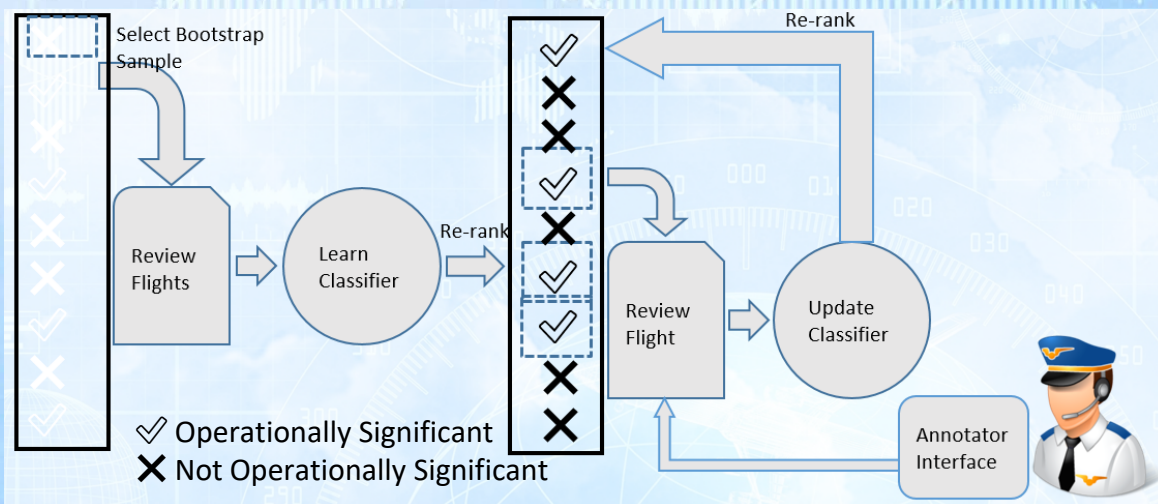
Anomaly Detection Pipeline



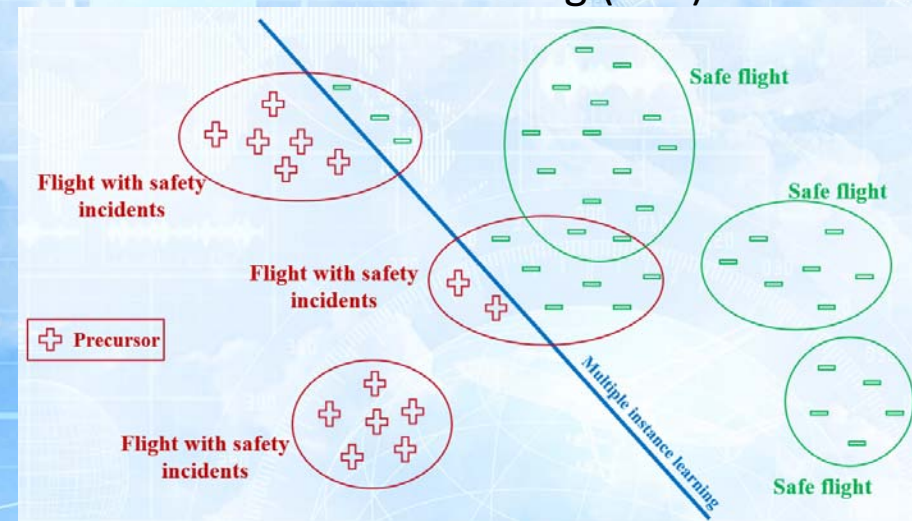
- IMS deviation scores need interpretation
- MMS post-processes IMS scores
- MMS score evaluates probability that each observation was generated from an off-nominal system

Anomaly Detection, cont.

Active Learning



Multi-instance Learning (MIL)



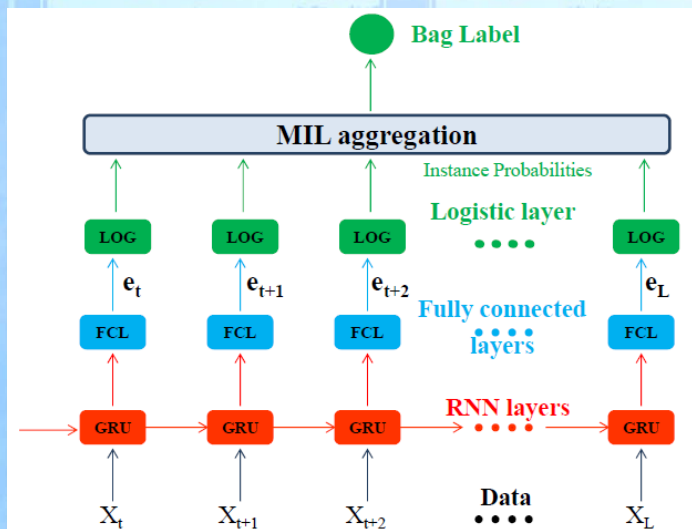
Testing

- 100 simulations of each failure type (rotor failure, tree collisions)
- Varied search area, number of drones, operating characteristics
- Train and test within scenarios, across scenarios



Findings:

- Successful detection of operationally significant vs. non-operationally significant anomalies
- Successful identification of precursors to anomalies

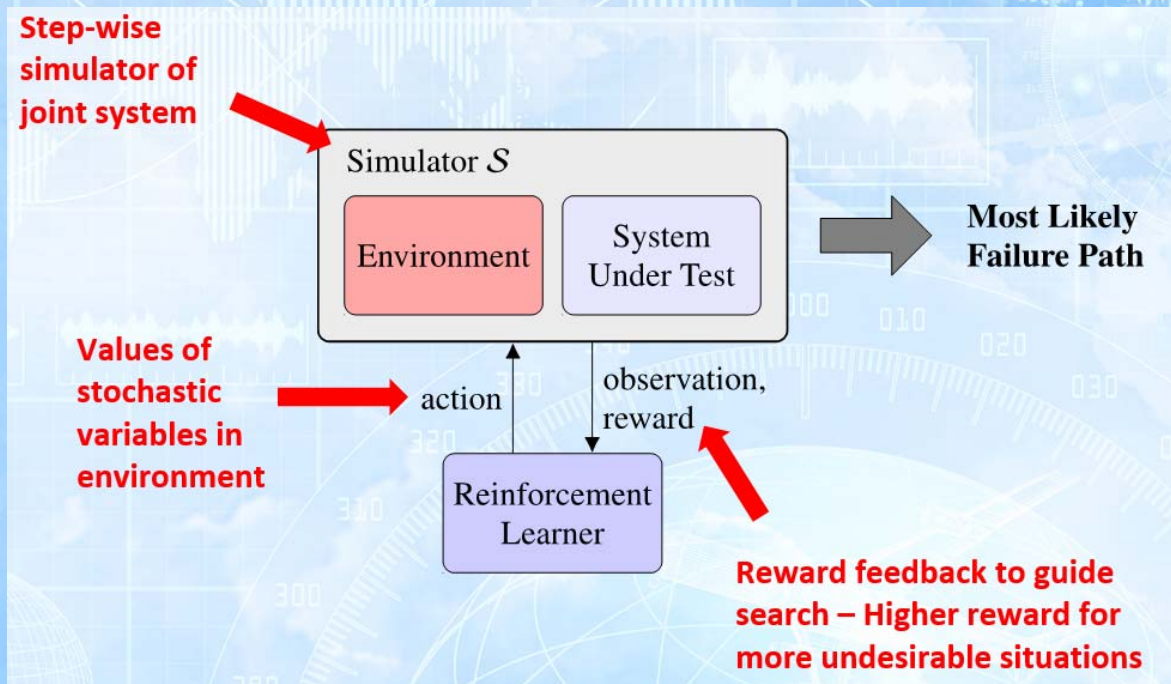




Adaptive Stress Testing of Trajectory Planning Systems

Ritchie Lee, Javier Puig-Navarro, Adrian Agogino, Dimitra Giannakopoulou, Ole Mengshoel, Mykel Kochenderfer, Danette Allen

- Idea/Concept: Validation of Trajectory Planners (TP) using Adaptive Stress Testing (AST) to increase confidence in TPs

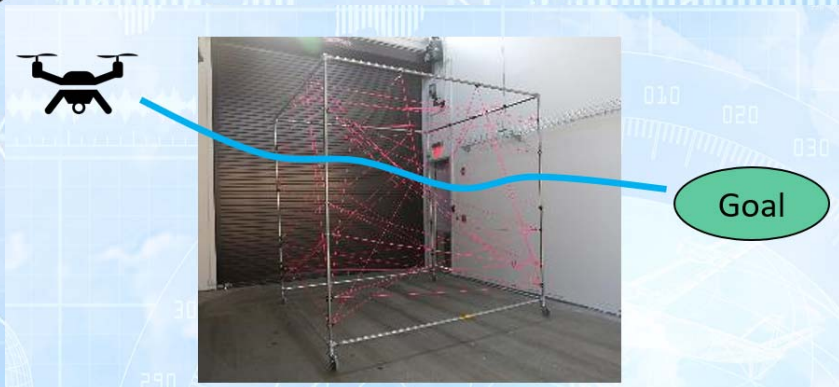


SOA (Monte-Carlo) does not scale

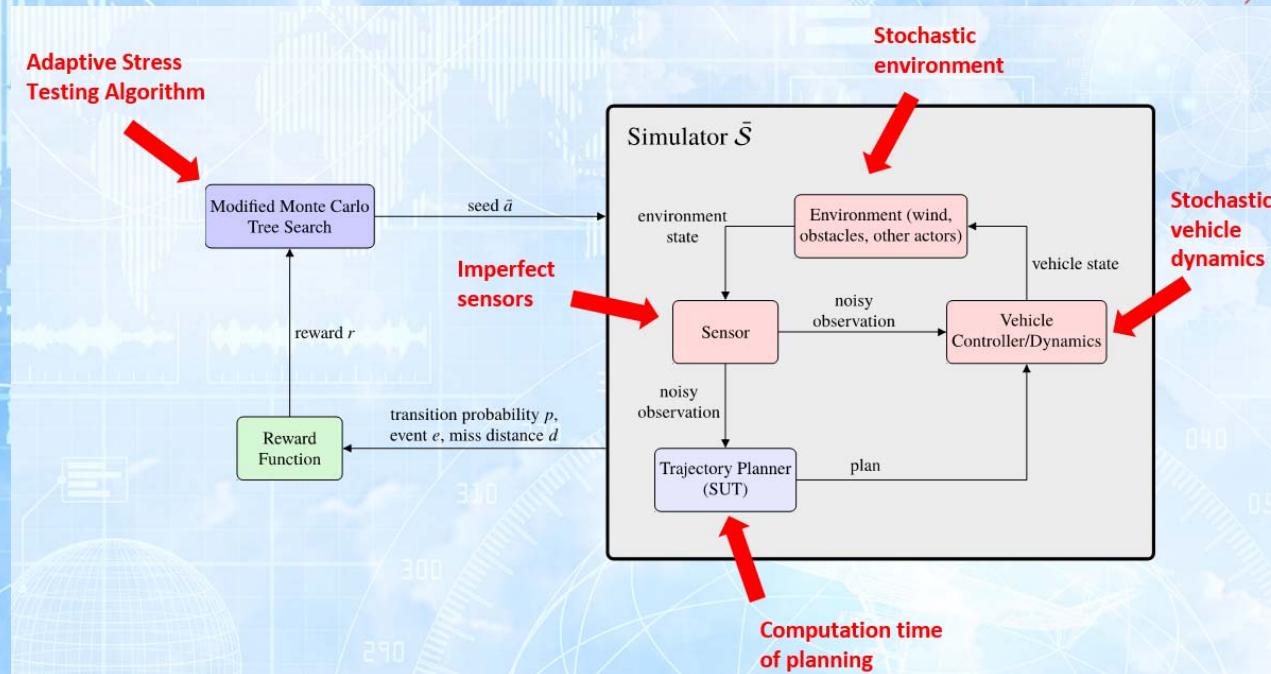
AST

- Formulates finding the **most likely failure path** as a **sequential decision-making problem**
- Apply reinforcement learning algorithms for finding failures
- Successful applications in Aircraft collision avoidance systems (ACAS X) and self-driving cars

AST: Application to Wire Maze Scenario



Testing Architecture



Experiments:

- Stressed for collision with obstacles; considered planning failures, implementation issues, etc.
- Testing dimensions: Sensor noise, initial configuration, replanning condition, computation time, dynamics noise

Results:

- 1600 total scenarios searched
- 386 successfully reached goal
- 19 collisions
- 1195 terminated in planning failure

Findings:

- AST can be extended and scaled to trajectory planning
- AST can run unanticipated scenarios and find a variety of errors
- Helped understand the residual risk, failure modes of the system
- Increased understanding of system's trustworthiness



Leveraging Fault-Tolerance Concepts as a Basis for Distributed Trust

Paul S. Miner

- [Idea/Concept](#): Justifiable *trust* is derived from a collection of *trustworthy monitors* that are coordinated using *trustworthy consensus mechanisms*
- Approach (work in progress): The working conjecture is that we can develop a workable model of trust by identifying/defining monitors and consensus mechanisms and determining how to ensure that the monitors and consensus mechanisms are sufficiently trustworthy for the level of trust required of the overall system
- Example of included attributes: Availability, Reliability, Safety, Confidentiality, Integrity, Maintainability; Relevant Knowledge; these are reasonably mature concepts but difficult to establish
- Hypothesis and Ongoing work: Well-established concepts from Dependable Computing apply to autonomous system and must be focused in specific contexts



Actionable Models of Trustworthiness and Trust



Natalia Alexandrov

- Idea/Concept: At the basis of every action in the system, at every scale, lies an explicit optimization problem. Consider T&T of optimization/decision making, individual and aggregated.

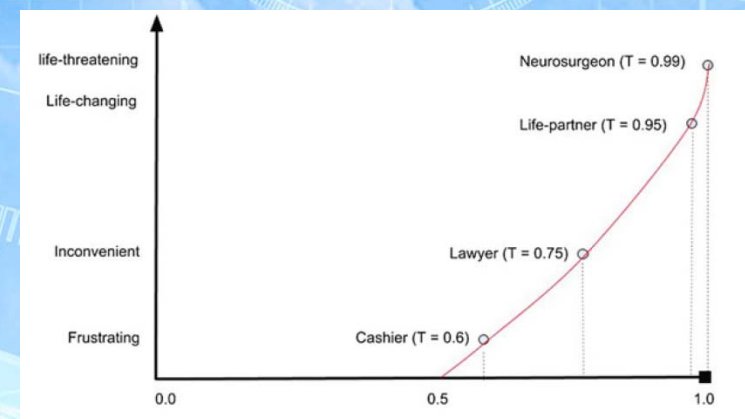
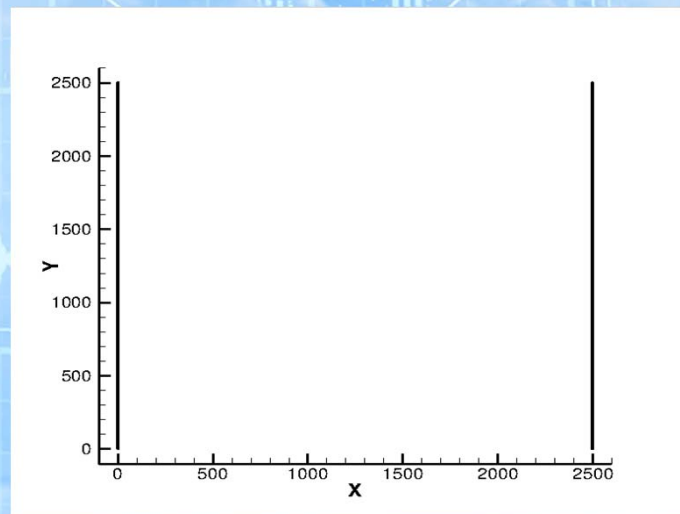
$$\text{Trustworthiness} \sim f(1/\text{Risk}, \text{thresholds})$$

- Trustworthiness: constraint for rational trust

- Agent A is trustworthy if:

- It has sufficient and correct information, sound decision-making processes, and shares mental models with other agents

- It exhibits accuracy, robustness, resilience, reliability, integrity, adaptability



Threshold for trust-based decisions, Abbas et al.

Ideal:

- Homogeneous operational aircraft
- Deterministic speed/position
- Sufficient time buffers
- Known intent and shared rules

Realistic:

- Heterogeneous aircraft of uncertain state
- Uncertain speed/position
- Uncertain time/space buffers
- Uncertain intent and rules

- Risk = $f(\text{uncertainty, impact of bad outcome})$
- Proposed approach to modeling the total epistemic uncertainty: Dempster-Shafer Evidence Theory
- Two components:

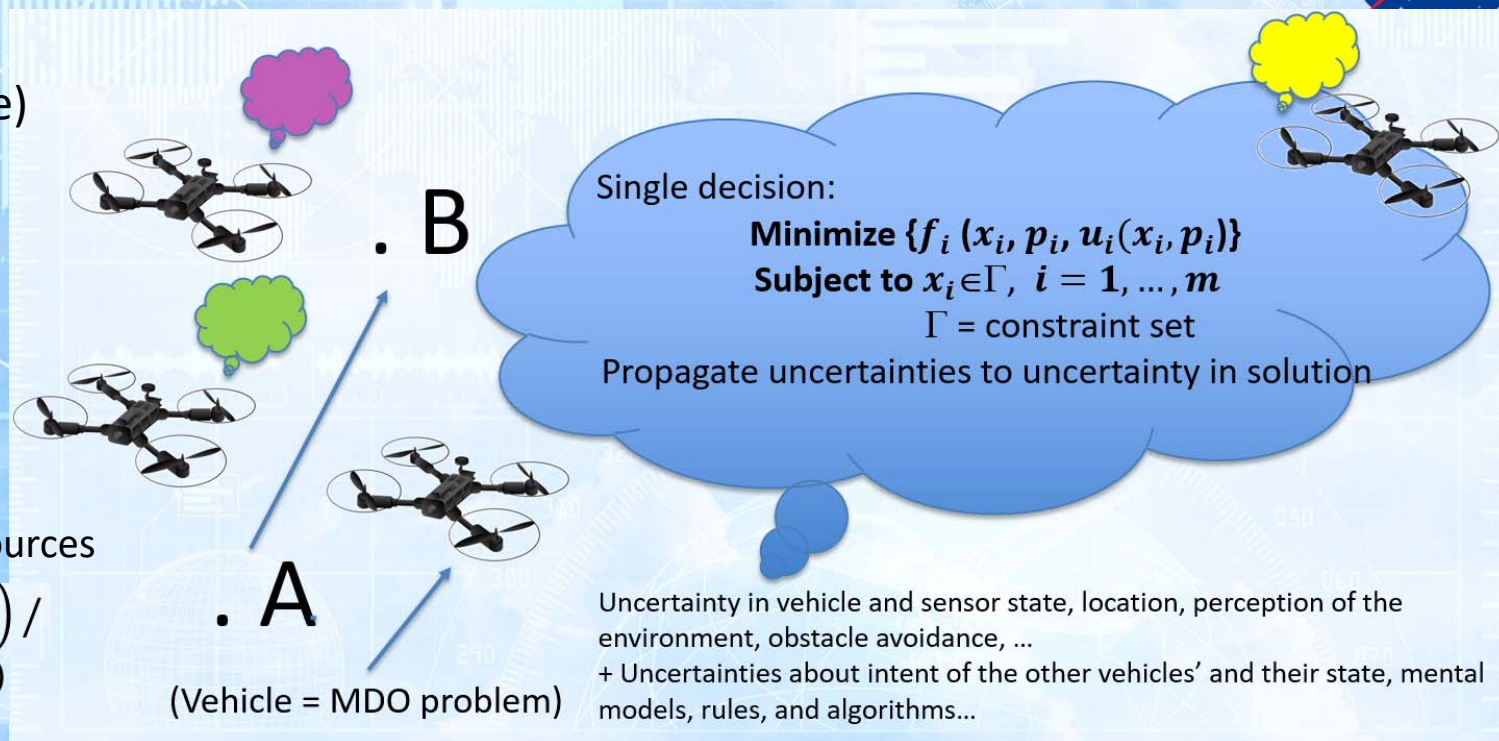
Representation of confidence in a proposition, due to evidence

Dempster's rule for combining evidence from various sources

$$m(C) = \left(\sum_{A_i \cap \dots \cap D_k = C} m_1(A_i) m_2(B_j) \dots m_n(D_k) \right) / \left(\sum_{A_j \cap \dots \cap D_k = \emptyset} m_1(A_i) m_2(B_j) \dots m_n(D_k) \right)$$

Findings:

- Optimization problem formulation and analysis, combined with evidence theory and logical-probabilistic representation is a feasible framework for explicit quantification of system's trustworthiness, for use in system design and real-time operations
- Practical usefulness depends of the quality of UQ, data, and models; development in specific context required
- Safety and performance of CPH are subject to irreducible uncertainties and must include externally imposed risk thresholds; robust/resilient design offers risk mitigation

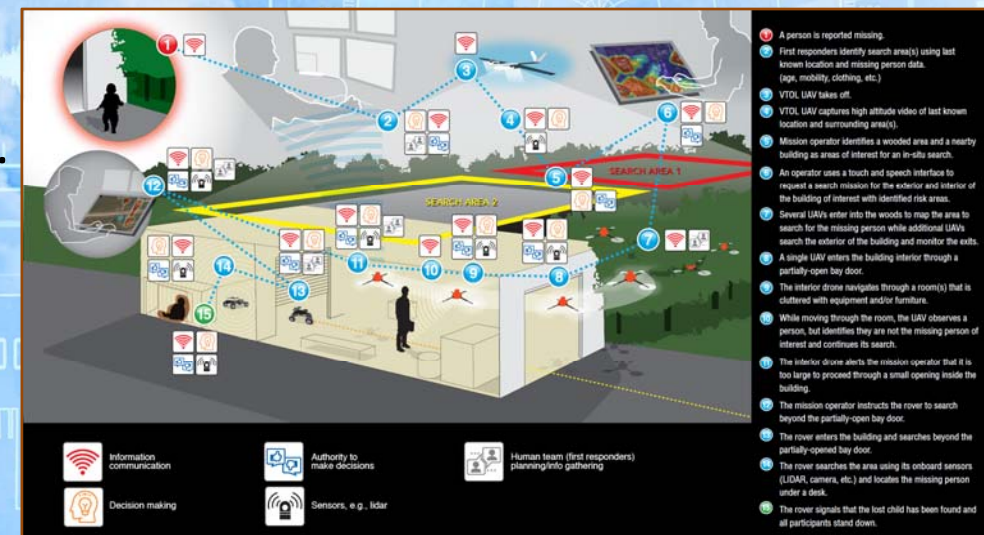


Cybersecurity Analysis

Ricaurte Chock, Kristen Hauser, Paul Nelson

Findings:

- The DRM is near-term feasible and achievable from a security perspective.
- The Protection Needs will not interfere with the project's goal.
- Opportunities afforded by DRM to test for various security concerns at scale “in the wild”:
 - How to field UAV, so that their information can be protected and tampering quickly identified as part of regular operations.
 - Authenticating and Authorizing UAV “in the wild” in an operational context in order to flexibly create trusted communication channels as needed during evolving operations.
- Trust issues and security principles such as those involved in the operational authentication and authorization are complex to solve in generalized environments such as the NAS.
- Area of concern: ML and its uses within the mission. This is due to the nature of ML.
- Projects such as ATTRACTOR needed to explore and provide a better understanding of how ML behaves and interacts within the overall system.



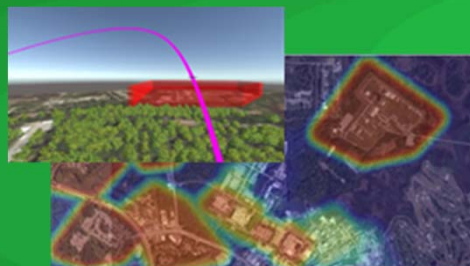


ATTRACTOR: Concluding Remarks

Multidisciplinary Component of
Complex Multi-agent Systems

Measurable trustworthiness
of fundamental functional
components

Context:
Trajectories
and DRM



Sound multi-objective
decision-making

Good Outcomes:
Basis for Trust

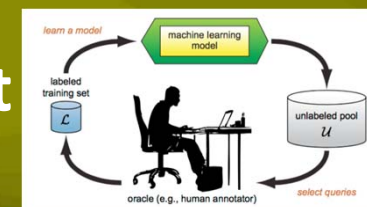
Heterogeneous multi-agent (including human-machine)
teaming is essential



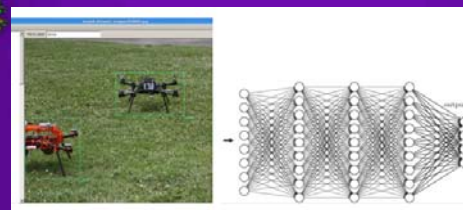
Teaming

Efficient two-way
learning/ training loop.

Learning &
Justifiable Trust



Explainability

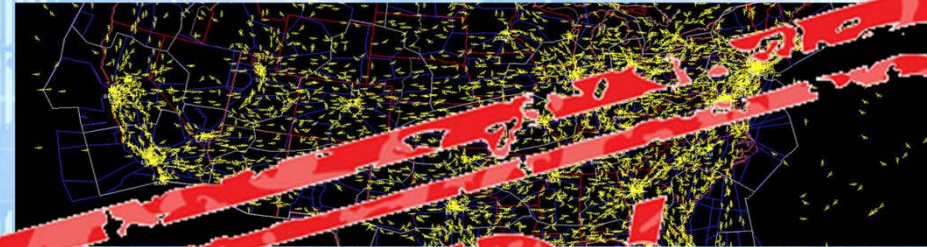


These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs.



Transition Status in FY20

- ARMD/AOSP/SWS
 - Adaptive stress testing
 - Actionable T&T modeling
 - Anomaly detection
- ARMD/AOSP/ATM-X
 - HMI/HAT
 - ML/AI
 - Modsim
- ARMD/TTT/AS
 - HMI/HAT
 - Modsim
- STMD/OSAM/OGA
 - Trajectory generation/management
 - Modsim
 - Multi-agent systems, swarms





Summary



- **Status:**

- Established either feasibility or clear direction of continued investigation into feasibility.
- The majority of capabilities have been integrated. Furlough of 2019 and COVID 2019 affected demo.
- Comparative measures of progress difficult as we are dealing with enabling, rather than improving
- Promising feedback from internal and external stakeholders (industry, OGA)
- V&V and T&T efforts are working to inform certification

- **Sample of future work directions:**

- Focus all development on specific application context
- Establishing safe operation envelopes for AI/ML
- Tractability of real-time computations during operations
- Explicit incorporation of context-relevant uncertainty quantification and risk for real-time decisions



Acknowledgments

Many thanks from the team to CAS for support of this work during execution and the bridge period.



Publications: Planting the AI Flag



2016–2019 PROGRESS REPORT:

ADVANCING ARTIFICIAL INTELLIGENCE R&D

From the Executive Office of the President of the United States

NASA investments in aeronautics have resulted in a new algorithm for active learning that uses limited domain expert feedback to reduce the false positive rate of data-driven anomaly-detection algorithms.⁹ NASA devised a new algorithm for automatically identifying precursors to known anomalies, where the precursors may happen significantly in advance of the anomalies. A project called Autonomy Teaming and TRAjectories for Complex Trusted Operational Reliability (ATTRACTOR) is developing a large set of technologies to enable explainable AI for multi-UAS (unmanned aircraft systems) missions during mission planning and execution.

2016–2019 PROGRESS REPORT: ADVANCING ARTIFICIAL INTELLIGENCE R&D

Within Earth science, NASA's Advanced Information Systems Technology (AIST) program (see sidebar) funds proposals from teams across NASA, universities, and industry for R&D and implementation of information technologies—including AI—for use in Earth science and downstream applications such as disaster response and water resource monitoring. Within aeronautics, NASA devises new algorithms for anomaly detection with the goal of automatically identifying anomalies within commercial aircraft flight-recorded data and radar track data. Within human space exploration and space science, integrated systems monitoring is needed to allow for health monitoring and maintenance of space systems during operations, due to the lack of human accessibility.

NASA	Developments for Earth and Beyond
NASA's AIST program has led to developments such as deep neural networks for land cover classification (e.g., asphalt, trees, grass, bare ground, or water); assessment of surface deformation due to earthquakes or volcanoes; assessment of coral reefs from multiresolution data through domain adaptation; symbolic regression to learn linear and nonlinear equations describing vegetation as a function of key climate variables; and time series search-and-analysis algorithms, including enhancements of these algorithms through quantum computing.	
NASA investments in aeronautics have resulted in a new algorithm for active learning that uses limited domain-expert feedback to reduce the false positive rate of data-driven anomaly-detection algorithms. ⁹ NASA devised a new algorithm for automatically identifying precursors to known anomalies, where the precursors may happen significantly in advance of the anomalies. A project called Autonomy Teaming and TRAjectories for Complex Trusted Operational Reliability (ATTRACTOR) is developing a large set of technologies to enable explainable AI for multi-UAS (unmanned aircraft systems) missions during mission planning and execution.	
NASA's human space exploration and space science activities have yielded clustering and deep learning ¹⁰ algorithms for monitoring and assessing the health of key systems such as the International Space Station and the Advanced Multi-Mission Operations System, which is a comprehensive system of tools to be used in future space missions. Automated planning and scheduling have been developed and used extensively for scheduling operations of systems such as the Mars rover, for scheduling crew operations, and for fault management. Machine learning is being developed as part of the project Autonomous Medical Operations for Deep Space Exploration, where the goal is development of medical models that are able to accurately diagnose a variety of pathologies that might occur on deep space missions, derived from ultrasound image analytics and/or machine learning.	

Active Anomaly Detection via Ensembles: Insights, Algorithms, and arxiv.org/abs/1901.08920.
of similar items in input data. Deep learning is an ML approach that can traditional programming or statistical methods. "Deep" refers to AI networks as opposed to the simpler networks studied earlier. <https://www.content.gov/2018/01/2018-National-Defense-Strategy-Update/277000089963/17-17/SUMMARY-OF-DDO-AI-STRATEGY-FDE>

– 6 –



2016–2019 PROGRESS REPORT:
ADVANCING ARTIFICIAL
INTELLIGENCE R&D

A report by the
ARTIFICIAL INTELLIGENCE RESEARCH & DEVELOPMENT
INTERAGENCY WORKING GROUP
SUBCOMMITTEE ON NETWORKING & INFORMATION TECHNOLOGY
RESEARCH & DEVELOPMENT
SUBCOMMITTEE ON MACHINE LEARNING & ARTIFICIAL
INTELLIGENCE
and the SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE
of the
NATIONAL SCIENCE & TECHNOLOGY COUNCIL
NOVEMBER 2019





Publications (Chronological Order)

- [0] Natalia Alexandrov and B Danette Allen, *Computational Models of Trustworthiness and Trust in Autonomous Cyber-Physical-Human Systems*, to be published at AIAA Aviation 2020.
- [0] B Danette Allen and Natalia Alexandrov, *Real-World and Mixed-Reality Applications of Trustworthiness and Trust in Autonomous Cyber-Physical-Human Systems*, to be published at AIAA Aviation 2020.
- [1] Bryan Barrows, Lisa Le Vie, Erica Meszaros, James Ecker, and B. D. Allen; *Analyzing Natural Language Context in Human-Machine Teaming using Supervised Machine Learning*; AIAA SciTech 2020
- [2] Philip C. Schulze, Justin P. Miller, David H. Klyde, Christopher D. Regan, Curtis Olson and Natalia Alexandrov; *Unmanned Aerial Systems Mission Task Element Development: Lessons Learned from Flight Test*, 6-10 January 2020, AIAA Scitech 2020 Forum
- [3] Ritchie Lee, J. Clarke, A. Agogino and D.Giannakopoulou: *Improving Trust in Deep Neural Networks with Nearest Neighbors*, AIAA 2020-2098, AIAA SciTech 2020
- [4] Ritchie Lee, Javier Puig-Navarro, Adrian K. Agogino, Dimitra Giannakoupoulou, Ole J. Mengshoel, Mykel J. Kochenderfer, and B. Danette Allen. *Adaptive Stress Testing of Trajectory Planning Systems*. In AIAA Scitech 2019 Forum, p. 1454. 2019
- [5] Javier Puig-Navarro, Naira Hovakimyan, and B Danette Allen: *Cooperative Trajectory Generation Framework for Cluttered Environments*, In AIAA Aviation 2019 Forum, p. 3254. 2019
- [6] N. Alexandrov and D. Allen; *ATTRACTOR: Building Trustworthiness and Trust in Autonomous Systems*, Aviation 2019, CAS Special Session
- [7] Chester V Dolph, Louis Glaab, Danette Allen, Maria Consiglio, and Khan Iftekharuddin; *An Improved Far-Field Small Unmanned Aerial System Optical Detection Algorithm*; DASC 2019
- [8] Meghan Chandarana, E. L. Meszaros, B. Danette Allen; *A Planning Pipeline for Large Multi-Agent Missions*, ACHI 2019
- [9] E. Meszaros, L. Le Vie, M. Last, B. Barrows, M. Smith, B. Allen; *Evaluating Communication Modality for Improved Human/Autonomous System Teaming*, AIAA SciTech 2019, AIAA-2019-2204
- [10] P. Chase Schulze, Justin P. Miller, David H. Klyde, Christopher D. Regan, and Natalia Alexandrov, *System Identification of a Small UAS in Support of Handling Qualities Evaluations*, AIAA 2019-0826, SciTech 2019
- [11] N. Alexandrov: *An Optimization Perspective on Trustworthiness and Trust in Autonomous Systems*, MOPTA 2019, 14-16 August 2019
- [12] Guynn, M. D.; Mukhopadhyay, V.; Mukhopadhyay, V.; Olson, E. D.; Ozoroski, T. A.; Alexandrov, N.; Olson, E. D.; Li, W.; and Kirk, J. T.: *Structural Analysis of Test Flight Vehicles with Multifunctional Energy Storage*. Presented at 2019 AIAA SciTech Forum and Exposition, January 7-11, 2019, San Diego, California



Publications (Chronological Order)

- [13] Alexander Grushin, Jyotirmaya Nanda, Ankit Tyagi, David Miller, Joshua Gluck, Nikunj C. Oza and Apoorv Maheshwari; *Decoding the Black Box: Extracting Explainable Decision Boundary Approximations from Machine Learning Models for Real Time Safety Assurance of the National Airspace*, AIAA-2019-0136, AIAA SciTech Forum 2019
- [14] Joshua Gluck, Ankit Tyagi, Alexander Grushin, David Miller, Sergey Voronin, Jyotirmaya Nanda and Nikunj C. Oza; *Too Fast, Too Low, and Too Close: Improved Real Time Safety Assurance of the National Airspace Using Long Short Term Memory*, AIAA-2019-0400, AIAA SciTech Forum
- [15] B. D. Allen; *Serious Gaming for Building a Basis of Certification for Trust and Trustworthiness of Autonomous Systems*, AIAA Aviation 2018
- [16] J. Puig - Navarro; N. Hovakimyan; N. Alexandrov; B. D. Allen; *Silhouette-Informed Trajectory Generation through a Wire Maze for UAS*, AIAA Aviation 2018
- [17] N. Alexandrov; *Trust, Trustworthiness and Risk in Autonomous Decision Making*, AIAA Aviation 2018
- [18] Paul S. Miner; *Leveraging Fault-Tolerance Concepts as a Basis for Distributed Trust*, AIAA Aviation 2018
- [19] David H. Klyde, Philip C. Schulze, David G. Mitchell, Natalia Alexandrov; *Development of a Process to Define Unmanned Aircraft Systems Handling Qualities*, (AIAA 2018-0299), 2018 AIAA Atmospheric Flight Mechanics Conference
- [20] Kamalika Das; David L. Iverson; Vijay Janakiraman; Nikunj Oza; *Incorporating Human Knowledge in Autonomous Systems through Machine Learning*, AIAA Aviation 2018
- [21] M. Chandarana; M. Lewis; K. Sycara; B. D. Allen; S. Scherer; *Swarm Size Planning Tool for Multi-Job Type Missions*, AIAA Aviation 2018
- [22] L. Tran, C. V. Dolph, and B. D. Allen; *Towards Explainability of UAV-Based Convolutional Neural Networks for Object Classification*, AIAA Aviation 2018
- [23] L. R. Le Vie; M. C. Last, B. A. Barrows, B. D. Allen; *Towards Informing an Intuitive Mission Planning Interface for Autonomous Multi-Asset Teams via Image Descriptions*, AIAA Aviation 2018
- [24] E L Meszaros; L R Le Vie; B D Allen; *Trusted Communication: Utilizing Speech Communication to Enhance Human-Machine Teaming Success*, Aviation 2018
- [25] B. N. Kelley; R. A. Williams; J. L. Holland; O. C. Schnarr; B. D. Allen; *A Persistent Simulation Environment for Autonomous Systems*, AIAA Aviation 2018
- [26] Yulun Tian, Katherine Liu, Kyel Ok, Loc Tran, Danette Allen, Nicholas Roy, and Jonathan How; *Search and Rescue under the Forest Canopy using Multiple UAS*, *International Symposium on Experimental Robotics* (ISER 2018), 140-152
- [27] Ecker, James and B Danette Allen, *Goal Detection via Mental Representation*, AIAA Aviation 2018
- [28] N. Alexandrov; *Explainable AI Decisions for Human-Autonomy Interactions*, 17th AIAA Aviation Technology, Integration, and Operations Conference, AIAA AVIATION Forum, AIAA 2017-3991

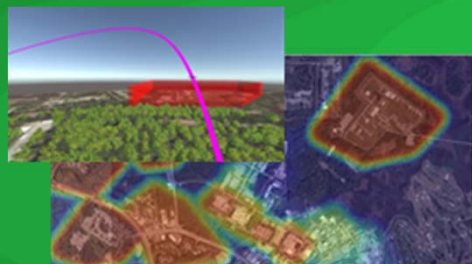


ATTRACTOR: Questions?

Multidisciplinary Component of
Complex Multi-agent Systems

Measurable trustworthiness
of fundamental functional
components

Context:
Trajectories
and DRM



Sound multi-objective
decision-making

Good Outcomes:
Basis for Trust

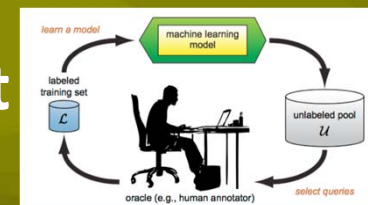
Heterogeneous multi-agent (including human-machine)
teaming is essential



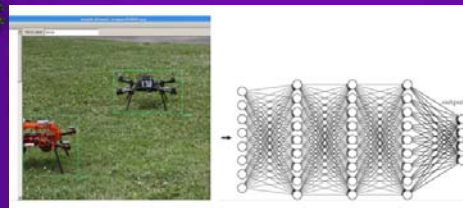
Teaming

Efficient two-way
learning/ training loop.

**Learning &
Justifiable Trust**



Explainability



These are two UAVs. Because...
They are quadrotors. They have four
propellers. They have four legs.