Prevailing in a Complex World: ARL's Essential Research Area on AI & ML

IST-160 Specialists' Meeting: Big Data & AI for Military Decision Making
30 May 2018

Tien Pham, Latasha Solomon, Greg Ciricionne, Brian Henz
ARL Computational & Information Sciences Directorate (CISD)

Topics

- Background Information
- AI & ML Essential Research Area
- Internal Research Efforts
- Collaborative Research Activities

Mission

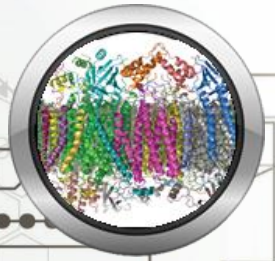
DISCOVER, INNOVATE, and TRANSITION Science and Technology to ensure dominant strategic land power

Computational Sciences

Analysis & Assessment

Information Sciences

Extramural Basic Research



Vision

The Nation's Premier Laboratory for Land Forces.

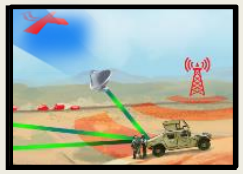
U.S. ARMY RDECOM®

ARL Essential Research Areas

ARL

Human-Agent Teaming

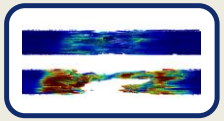
Cyber & EM Technologies for Complex Environments



Distributed / Cooperative Engagement in Contested Environments

Artificial Intelligence/ Machine Learning

Manipulate Failure Physics for Robust Materials



Tactical Unit Energy Independence

Accelerated Learning for a Ready Force

Manufacturing at the Point of Need

Discovery

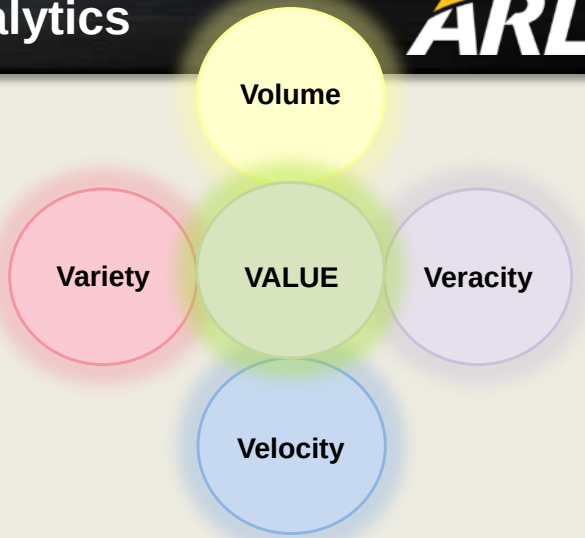
ERA's aligned to the Chief of Staff Army's Modernizations Priorities

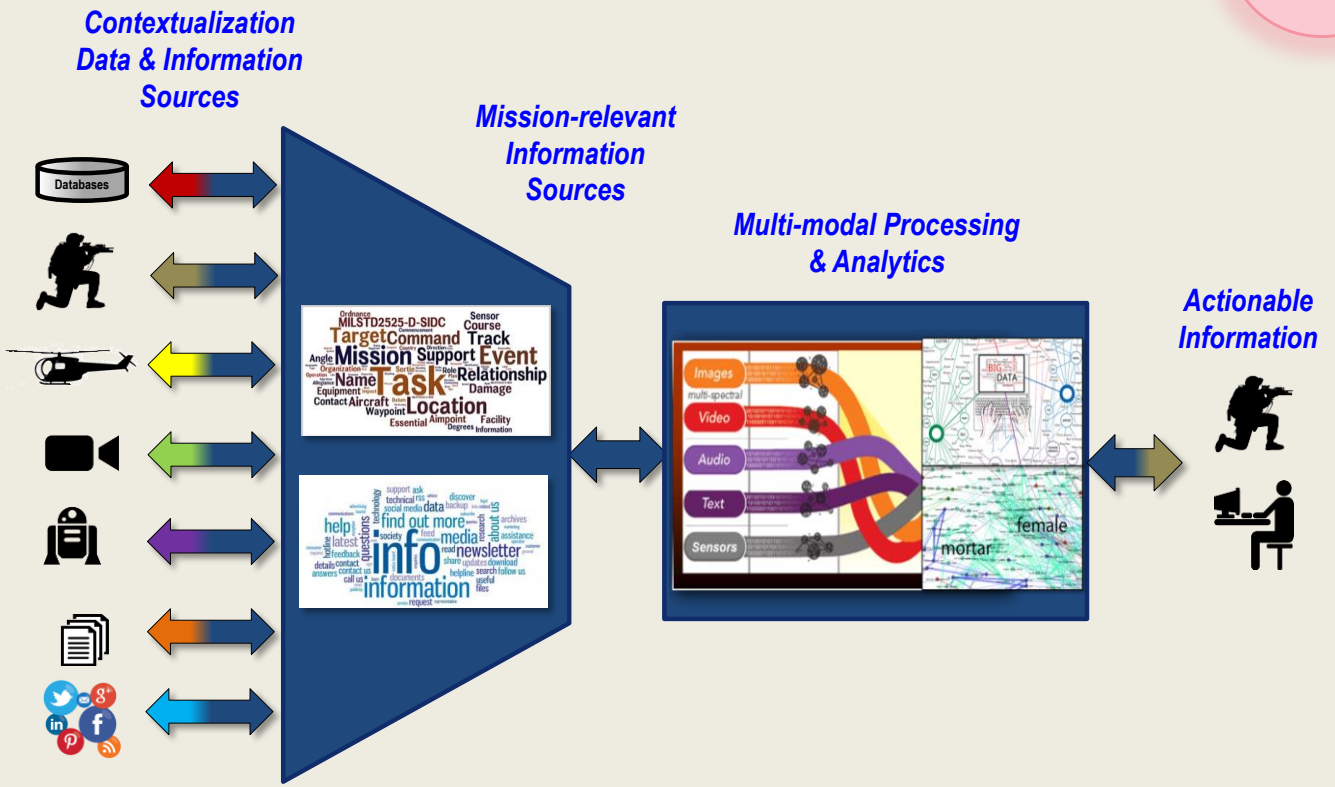
1. Long Range Precision Fires
2. Next Gen Vehicle
3. Future Vertical Lift
4. Networks/C3I
5. Air & Missile Defense
6. Soldier Lethality

The Nation's Premier Laboratory for Land Forces

Big Data Analytics

- **5 V's of Big Data: Volume** (vast amount), **Velocity** (speed of data generation), **Variety** (diversity of data types), **Veracity** (messiness and trustworthiness) and **Value** (usefulness)





Big Data = "Distributed, Disparate, Dinky, Dirty, Dynamic, Deceptive Data" (D7)

Topics

- Background Information
- AI & ML Essential Research Area
- Internal Research Efforts
- Collaborative Research Activities

Multi Domain Battle → Prevailing in a Complex World

Large-scale, cluttered, contested urban environment

Highly-dispersed team of human & robot agents accessing highly heterogeneous information sources

Dynamic in-flight learning & re-planning at the Speed of the Fight

Learning in new environments with deception from persistent threats

Focused research to address CSA Priorities: (i) Next Gen Combat Vehicles and (ii) Networks/C3I

AI & ML ERA

Human-Agent Teaming
2. Next Gen Combat Vehicle

Cyber & EM Technologies for Complex Environments
4. Network/C3I

Artificial Intelligence/ Machine Learning
2. Next Gen Combat Vehicle
4. Networks/C3I

Multi-Domain Battle (MDB) *describes how U.S. ground forces, as part of the Joint Force and with partners, will operate, fight, and campaign successfully across all domains—space, cyberspace, air, land, maritime—against peer adversaries in the 2025-2040 timeframe.*

http://www.arcic.army.mil/App_Documents/Multi-Domain-Battle-Evolution-of-Combined-Arms.pdf

- **AI & ML ERA addresses Army-relevant research gaps in MDB**
- **Focusing AI & ML research within the context of 2 other ERA's**
  - **HAT: Highly dispersed team of humans and agents (robots and software)**
  - **CETCE: operating in congested and contest environment with near-peer adversaries**

AI & ML Research Challenges

AI & ML Research Gaps

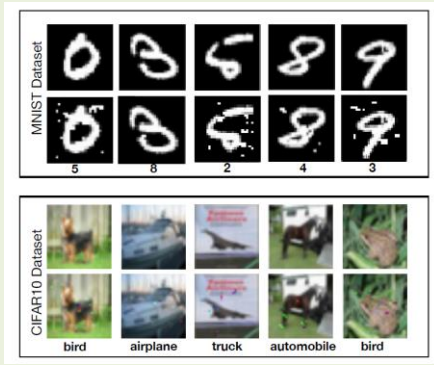
Learning in Complex Data Environments	→ AI & ML with small samples, dirty data, high clutter → AI & ML with highly heterogeneous data → Adversarial AI & ML in contested, deceptive environment
Resource-constrained AI Processing at the Point-of-Need	→ Distributed AI & ML with limited communications → AI & ML computing with extremely low size, weight, and power, time available (SWaPT)
Generalizable & Predictable AI	→ Explainability & programmability for AI & ML → AI & ML with integrated quantitative models

Goal: **To research and develop artificially intelligent agents (heterogeneous & distributed) that rapidly learn, adapt, reason & act in contested, austere & congested environments**

Adversarial Machine Learning

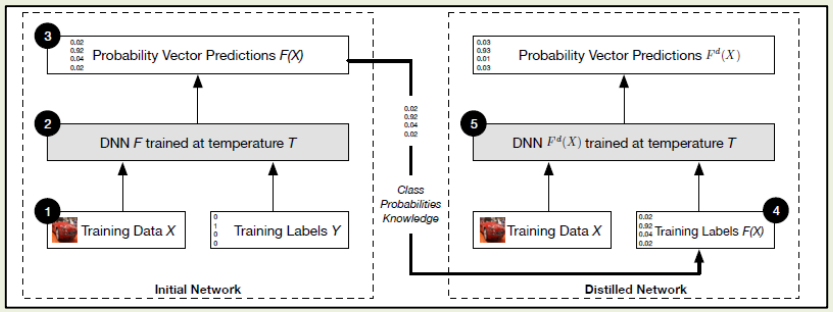
Innovation: Characterized vulnerability of Deep Neural Networks (DNNs) arising from training imperfection; introduced novel class of algorithms to craft adversarial samples and demonstrated cross-model transferability of adversarial samples; developed a defensive mechanism – distillation - to reduce effectiveness of adversarial samples.

Key Technical Demonstration:



Gradient-based & Gradient-free Adversarial Sample Generation

Defensive Distillation



Benchmarking ML Algorithms
Open-source code: ARL Cyber CRA
github.com/tensorflow/cleverhans

DAIS ITA: (Audio)
https://git.io/vFs8x
https://git.io/vFs42

2-orders of magnitude reduction in vulnerability to adversarial samples in proof-of-concept system & in adversarial sample generation.

Robust ML in dynamic, distributed, deceptive & contested environment

Key Publications:

Alzantot et al. Did you hear that? Adversarial examples against automatic speech recognition. NIPS 2017.

Grosse et al. *Adversarial Perturbations Against Deep Neural Networks for Malware Classification.* ESORICS 2017.

Papernot et al. *Practical black-box attacks against ML.* AsiaCCS 2017.

Celik et al (2018). *Detection under privileged information.* AsiaCCS 2018

Alzantot et al. *GenAttack: Practical Black-box Attacks with Gradient-Free Optimization (submitted to ICML 2018).*

Alzantot et al. Generating Natural Language Adversarial Samples *(sub to ACL 2018)*

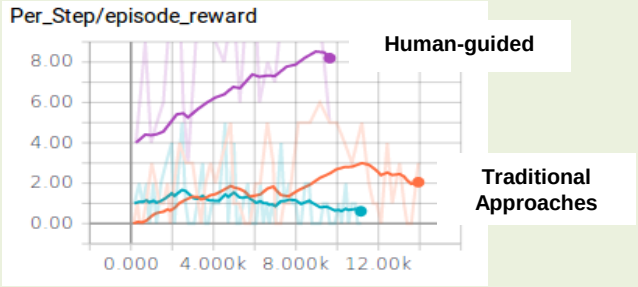
Potential:

- Robust training of ML algorithms against adversarial attacks
- Rapid training with fewer samples: image, video, speech, text
- Quantification of complexity-accuracy-resilience tradeoffs

AI Learning from Humans

Innovation: Human-in-the-loop reinforcement learning system to provide improve decision-making in *dynamically-changing environments*, where data availability and computational resources are limited.

Key Technical Demonstration: **Reinforcement-learning AI solution using real-time human input to solve the unsolved Atari™ Bowling task**



Philosophical similarities with OpenAI/DeepMind research, but with ability to run in real-time and claimed significant improvements in sample efficiency - Jack Clark, Director of Strategy and Communications at Open AI.

- **Converged on state-of-the-art solution for previously unsolved Atari™ Bowling task in 15 minutes**
- **Learned AI policy outperforms expert humans**

Reconceiving human-technology roles in the future Battlefield

Key Publications:

Warnell et al. (submitted) *Deep TAMER: Interactive Agent Shaping in High Dimensional State Spaces*. AAAI-18.
Koppel et al (submitted) *Policy Eval. in Infinite MDPs: Eff. Kernel Gradient Temporal Difference*. AAAI-18.

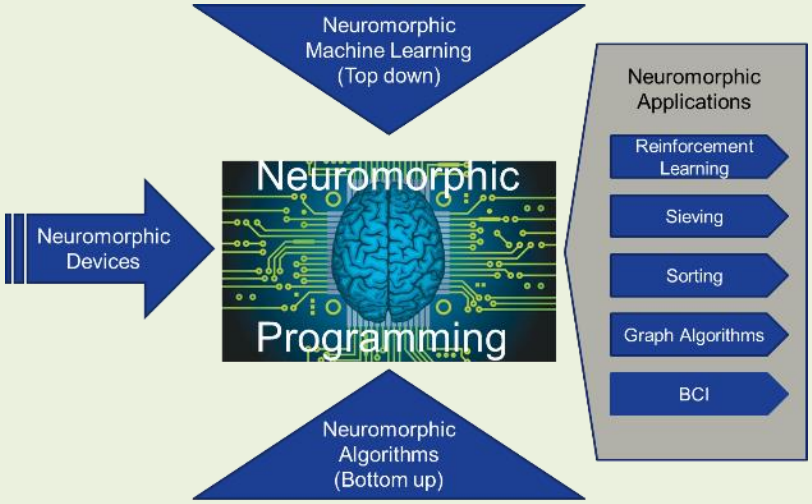
Potential:

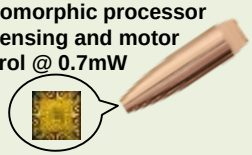
- Broader applications of AI through solutions for unstructured environments with ill-defined rewards.
- Faster, more optimal AI solutions with less data
- Rapidly-adaptable Human-AI teams that learn from human understanding of high-level goals

AI Processing At-the-Point of Need

Innovation: Developed a framework for creating, deploying, and executing tightly-integrated symbolic and gradient-based learned neural networks on neuromorphic systems

Key Technical Demonstration: End-to-end integration of systems programming and deep learning run entirely on neuromorphic devices





- **Accelerated computing and AI systems that run in real-time at less than ½ Watt of power**

Seamlessly composable symbolic processing and gradient-based learning in a single neural network

Key Publications:

Dawson et al (2017) *Tightly integrated deep learning and symbolic prog. on a single neuromorphic chip. 2017 IEEE Int Sym on Circ & Sys (ISCAS)*

Potential:

- Embed processing for AI systems integrated in a broad range of devices with extremely low size, weight, power, and processing time requirements

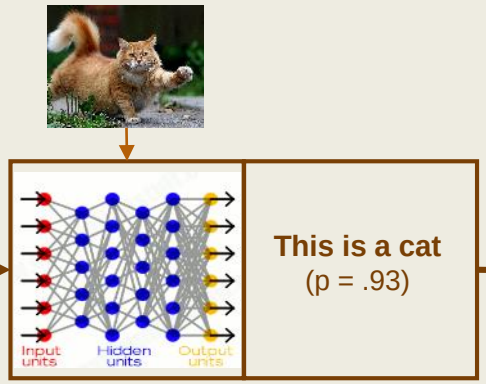
Explainable AI (XAI)

XAI
EXPLAINABLE ARTIFICIAL INTELLIGENCE

Today

Training Data

Learning Process



Learned Function

This is a cat
(p = .93)

Output

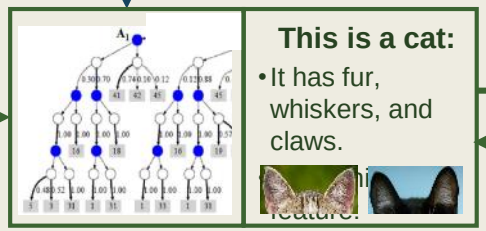
User with a Task

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Tomorrow

Training Data

New Learning Process



Explainable Model

This is a cat:
- It has fur, whiskers, and claws.

Explanation Interface

User with a Task

- I understand why
- I understand why not
- I know when you'll succeed
- I know when you'll fail
- I know when to trust you
- I know why you erred

http://www.darpa.mil/program/explainable-artificial-intelligence

Courtesy: David Gunning, PM XAI

Topics

- Background Information
- AI & ML Essential Research Area
- **Internal Research Efforts**
- **Collaborative Research Activities**

AI-enabled Scenario - “Fight’s Eye”

A Soldier supported by team of agents in complex environment

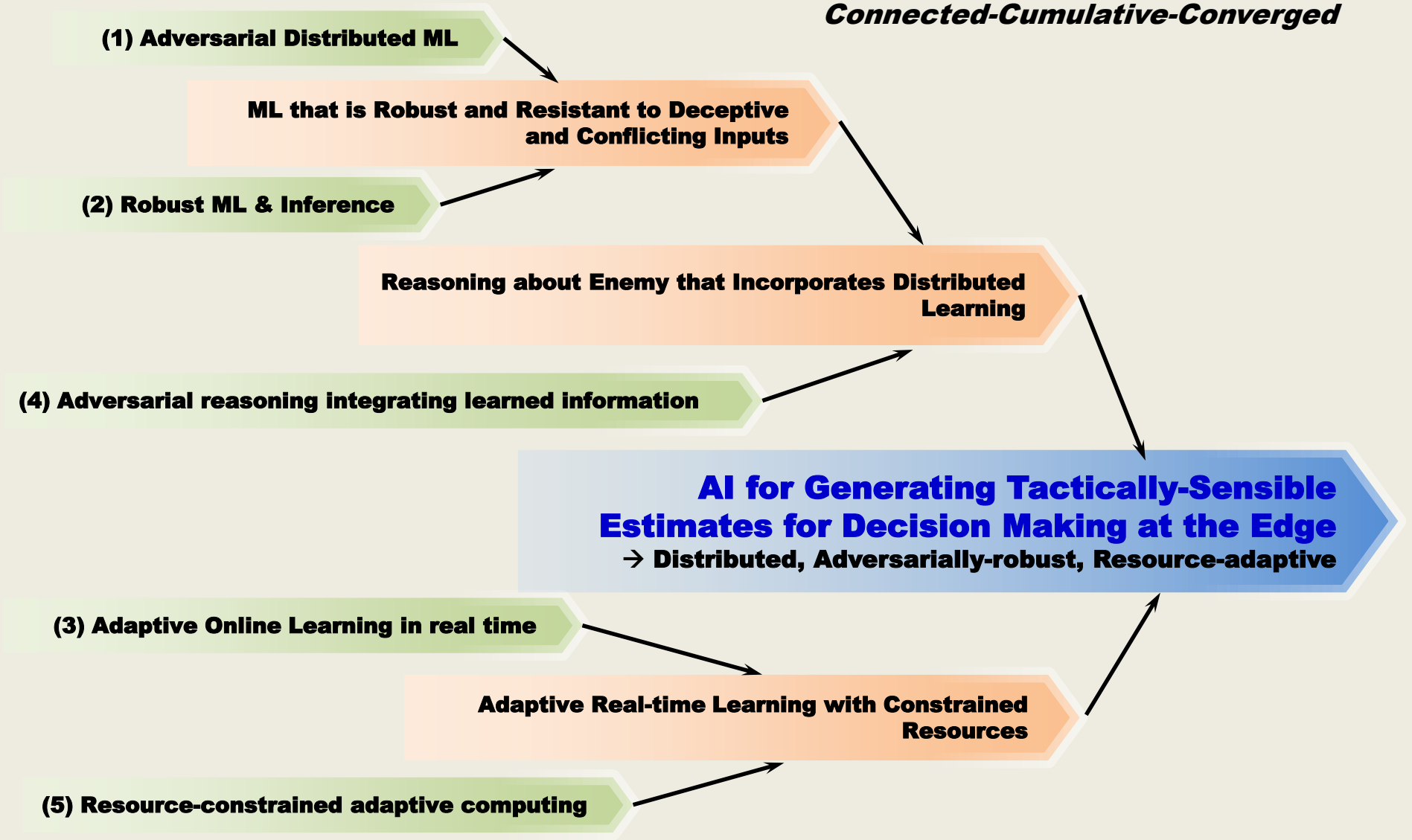
AI-enabled real-time estimates of enemy

AI-enabled reasoning to provide possible course of actions

Tactically sensible decision making based on locally available information

Conceptual Roadmap

Connected-Cumulative-Converged



1. Adversarial Distributed ML

- Characterization of vulnerabilities in ML, development of defenses
- Distributed Learning: heterogeneity of models, fusing models, under uncertainty & deception
- Interpretable Learning
- Scenario development and joint experimentation

2. Robust ML and Inference

- Characterization and quantitative formulation of type of uncertainty
- Learning and inference over sparse multimodal data
- Baseline and agile reasoning leveraging machine learning and human expertise

3. Adaptive Online Learning

- Learning for high-speed navigation in unknown environments
- Online unsupervised percept modeling
- Stable and risk-aware learning and adaptation

4. Adversarial Reasoning

- Information subspace reduction
- Uncertainty quantification / estimation and logic representation
- Quantitative causality and adversarial discovery

5. Resource-constrained Adaptive Computing

- Techniques for adaptive allocation of computing resources to tasks
- Real-time adaptation of algorithms to computational tasks of various complexity
- Methods for optimized real-time reconfiguration of hardware based on properties of the tasks and available software

Topics

- Background Information
- AI & ML Essential Research Area
- Internal Research Efforts
- **Collaborative Research Activities**

Collaborative Research Programs

Network Science Collaborative Technology Alliance

NETWORK SCIENCE CTA

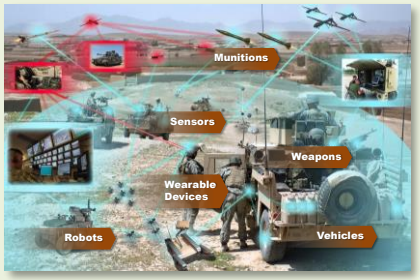
Composite Networks: Understand, Predict, and Influence

Ends in Sep 2019

Cyber-Security Collaborative Research Alliance

Fundamental understanding of cyber phenomena

Internet of Battlefield Things (IoBT) CRA



Started in Oct 2017

Robotics Collaborative Technology Alliance

RCTA Robotics Collaborative Technology Alliance

Enable the creation of future intelligent unmanned systems

Ends in Sep 2019

Distributed Analytics & Information Sciences ITA

IBM, Raytheon, BAE SYSTEMS, AIRBUS, UCLA, DAIS ITA, ARL, [dstl]

Started in Sep 2016

Distributed Collaborative Intelligent Systems (DCIST) CRA

Started in Oct 2017

ARL collaborative research programs with external academic & industrial partners address **Big Data challenges** in **distributed analytics** and **IoT**, **networks** and **cyber security**, **autonomy** and **robotics**

Flagship program with an alliance of U.S. & UK gov't, industrial & academic researchers developing underpinnings of DAIS to enhance capabilities to conduct coalition operations

Research Program

P1: Software Defined Coalitions

P2: Generative Policy Models for Coalitions

P3: Agile Composition for Coalition Environments

P4: Instinctive Analytics in a Coalition Environment

T1: Resource Allocation for Dynamically Formed Distributed Analytics Services

T2: Self-aware Cognitive Services for Distributed Coalition Environments

P5: Anticipatory Situational Understanding for Coalitions

T1: Learning and Reasoning in Complex Coalition Information Environments

T2: Interpretable Deep Neural Networks for Coalition Situational Understanding

P6: Evolution of Complex Adaptive Human Systems

Internet of Battlefield Things (IoBT) CRA

An IoBT is a *set* of interdependent entities or things

- Sensors, actuators, devices
- Infrastructure
- Analytics
- Information Sources & Open Source Intelligence
- Humans

Research Areas (RAs)

- **RA1**: Discovery, Composition and Adaptation of Goal-Driven Heterogeneous IoBTs
- **RA2**: Autonomic IoBTs to Enable Intelligent Services
- **RA3**: Distributed Asynchronous Processing and Analytics of Things
- **CCRI**: Cyber-Physical Security

Consortium Members: UIUC (lead), UMass, UCLA, USC, CMU, UC Berkeley, SRI International

DCIST Vision

- Highly distributed and collaborative heterogeneous teams of intelligent systems
- Integrate varying levels of autonomy and intelligence with the Soldier
- Augment the capability of the collective
  well beyond that of any one component

Consortium Members: U Penn (lead), MIT, Georgia Tech, UCSD, UC Berkeley, USC

Research Areas (RAs)

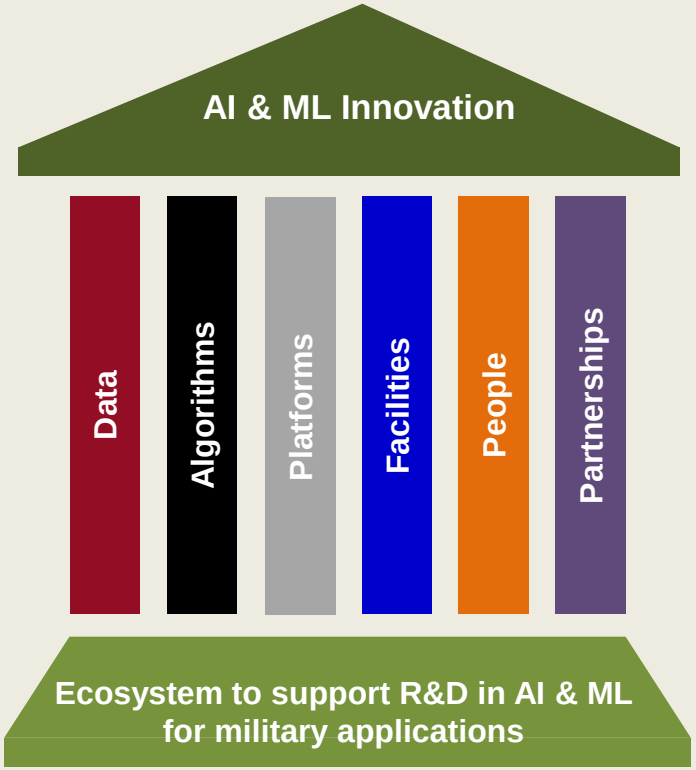
- **RA1**: Distributed Intelligence
- **RA2**: Heterogeneous Group Control
- **RA3**: Adaptive and Resilient Behaviors

Prevailing in a Complex World: Essential Research Area on AI & ML

IST-160 Specialists' Meeting: Big Data & AI for Military Decision Making
Panelists' Discussion – 1 June, 2018

Tien Pham
Information Sciences Campaign

Enabling AI & ML Research



- **Goal of AI & ML research is to rapidly advance Adaptive and Robust AI & ML capabilities for *Multi Domain Battle***
- **Center/hub/forum to serve as a focal point for AI & ML research and innovation**
  - Collate existing data, collect multi-domain data and make data accessible
  - Develop complex scenarios, use cases and challenge problems with associated data
  - Provide computing resources, software tools and V&V capabilities
  - Perform experiments and demonstrations in complex environments
  - Transition new, advanced, robust AI & ML algorithms
  - Educate military leaders and end users

AI & ML research needs: (i) data (multi domain), (ii) computing resources (HPC and edge computing platforms), and (iii) algorithms (for complex environment)

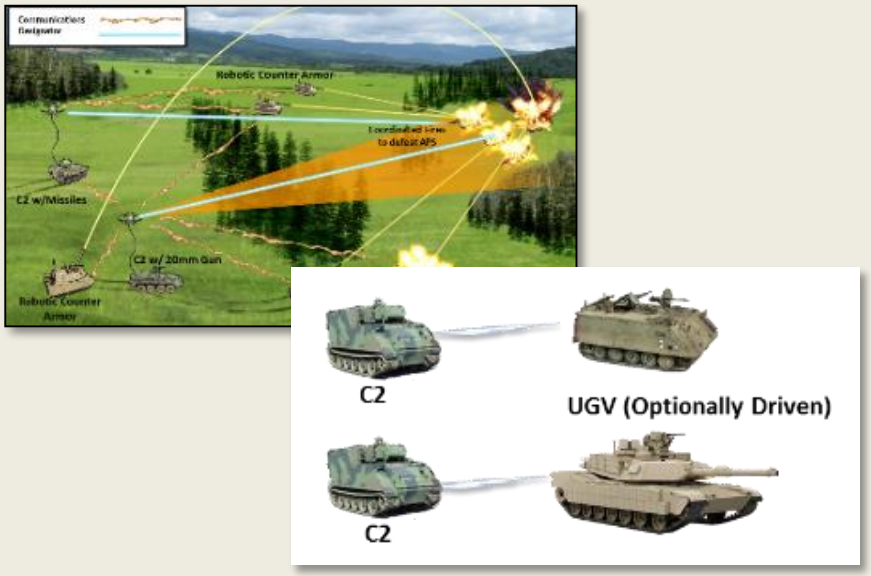
Example: Data & Challenge Problem

Military Relevant Datasets for Autonomous Systems (planned)

- **Goal** is to create a large-scale, ground-truthed, multi-modal, military-relevant dataset
- **Focus Areas**: Sensor, development, data collection & ground, truth, data annotation, repository of data

Sample Realistic Environments

Autonomous Maneuver Challenge Problem



Objective: Advance the state of the art in autonomous ground maneuver in complex military environments

Phase 1 – Simulation

Phase 2 – On live robots in realistic environments

Phase 3 – Complex terrain as exemplified in Next Gen Combat Vehicle

Example: ML Algorithm Evaluation Tool

CleverHans Library for benchmarking ML systems against adversarial examples

Problem

- Reproducing research results on adversarial examples was difficult because **implementations of attacks and defenses were not standardized**

Approach

- Created the **CleverHans library**, in a joint effort with Ian Goodfellow at Google.
- Contains **reference implementations** of several attack and defense procedures.
- Researchers/product developers canuse cleverHans to **test their models against standardized, state-of-the-art attacks and defenses**.
- **Results in published research are comparable to one another**

Papernot et al. CleverHans v2.1.0: an adversarial machine learning library. arXiv:1610.00768v5

Results

- CleverHans is an **active open-source project** on **GitHub** (stats Apr 2018)
  - Created: October 2016
  - Stars: > 1700
  - Forks: > 450
  - Repository views per week: > 9K
  - Repository clones per week: > 450
  - Contributors: > 50 (made over 1,400 commits)
- Used in a **NIPS 2017 Kaggle competition:** 65 teams from a variety of universities across the globe

cleverhans

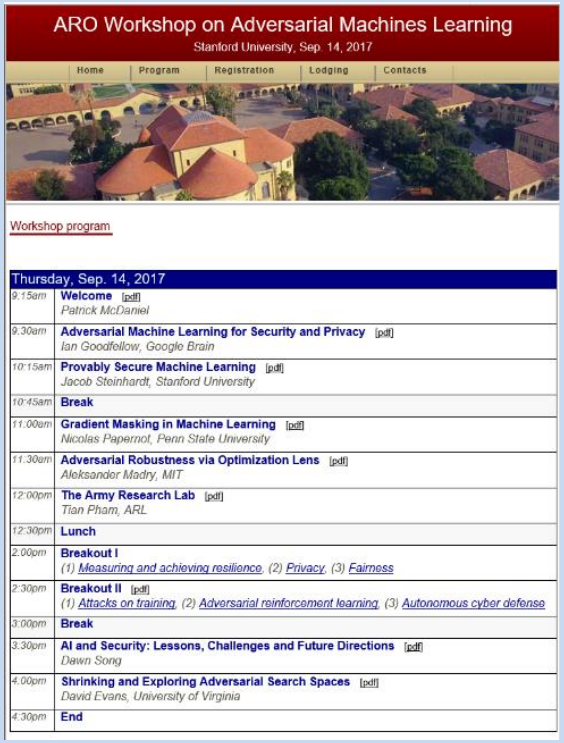
https://github.com/tensorflow/cleverhans

Example: ML Workshops

Adversarial Machine Learning #1

Stanford University – 14 September 2017

The workshop aims to bring together researchers working on adversarial machine learning.



ARO Workshop on Adversarial Machines Learning

Stanford University, Sep. 14, 2017

Home | Program | Registration | Lodging | Contacts

Workshop program

Thursday, Sep. 14, 2017	
9:15am	**Welcome** [pdf] *Patrick McDaniel*
9:30am	**Adversarial Machine Learning for Security and Privacy** [pdf] *Ian Goodfellow, Google Brain*
10:15am	**Provably Secure Machine Learning** [pdf] *Jacob Steinhardt, Stanford University*
10:45am	**Break**
11:00am	**Gradient Masking in Machine Learning** [pdf] *Nicolas Papernot, Penn State University*
11:30am	**Adversarial Robustness via Optimization Lens** [pdf] *Aleksander Madry, MIT*
12:00pm	**The Army Research Lab** [pdf] *Tian Pham, ARL*
12:30pm	**Lunch**
2:00pm	**Breakout I** (1) *Measuring and achieving resilience*, (2) *Privacy*, (3) *Fairness*
2:30pm	**Breakout II** [pdf] (1) *Attacks on training*, (2) *Adversarial reinforcement learning*, (3) *Autonomous cyber defense*
3:00pm	**Break**
3:30pm	**AI and Security: Lessons, Challenges and Future Directions** [pdf] *Dawn Song*
4:00pm	**Shrinking and Exploring Adversarial Search Spaces** [pdf] *David Evans, University of Virginia*
4:30pm	**End**

Adversarial Machine Learning #2

Joint ARL and IARPA Workshop
University of Maryland – 9 May 2018

- **11 invited talks and 100+ participants from academia, industry and government.**
- **Plan to convert workshop into a formal adversarial learning conference in 2019**

Organize workshops on specific AI technical topics.
Organize specifc AI courses for developers, end users and military leaders.

Artificial Intelligence and Machine Learning for Multi Domain Battle Applications

Conference Chair: Tien Pham (ARL)

Conference Co-Chairs: Guest Chair (OGA/Industry/Academia) and Latasha Solomon (ARL)

Program Committee: ARL, UK DSTL, AFOSR, SPAWAR, IBM US, IBM UK, Lockheed Martin, Discovery Lab Global, UIUC, Cardiff University, University of Calgary, …

Focus: AI/ML for the future Multi Domain Battle (MDB) operations that involve teams of highly-dispersed warfighters and agents (robotic and software) operating in distributed, dynamic, complex, cluttered environments.

Topics include but not limited to:

1: Learning in Complex Environment
- AI & ML with small samples, dirty data, high clutter
- AI & ML with highly heterogeneous data
- Adversarial AI & ML in contested, deceptive environment

2: Resource-constrained AI Processing at the Point-of-Need
- Distributed AI & ML with limited communications
- AI & ML computing with extremely SWaPT

3: Generalizable and Predictable AI
- Explainability & programmability for AI & ML
- AI & ML with integrated quantitative models

4: Human Agent Teaming
- Human-to-AI and AI-to-Human interaction and understanding
- Collaborative intelligence

5: Distributed Analytics for Situational Understanding

6: Internet of Things (IoT) for Battlefield Applications

7. Uncertainty Characterization and Quantification

April 2019, Baltimore, MD

SPIE. DCS DEFENSE+ SECURITY

Welcome additional international participants on program committee

Contact Information

Tien Pham, Ph.D.
Senior Campaign Scientist
Information Sciences

U.S. Army Research Laboratory
Attn: RDRL-CI
2800 Powder Mill Road
Adelphi, MD 20783 USA
Tel: +1-301-394-4282
Mobile: +1-240-997-2768
Email: tien.pham1.civ@mail.mil