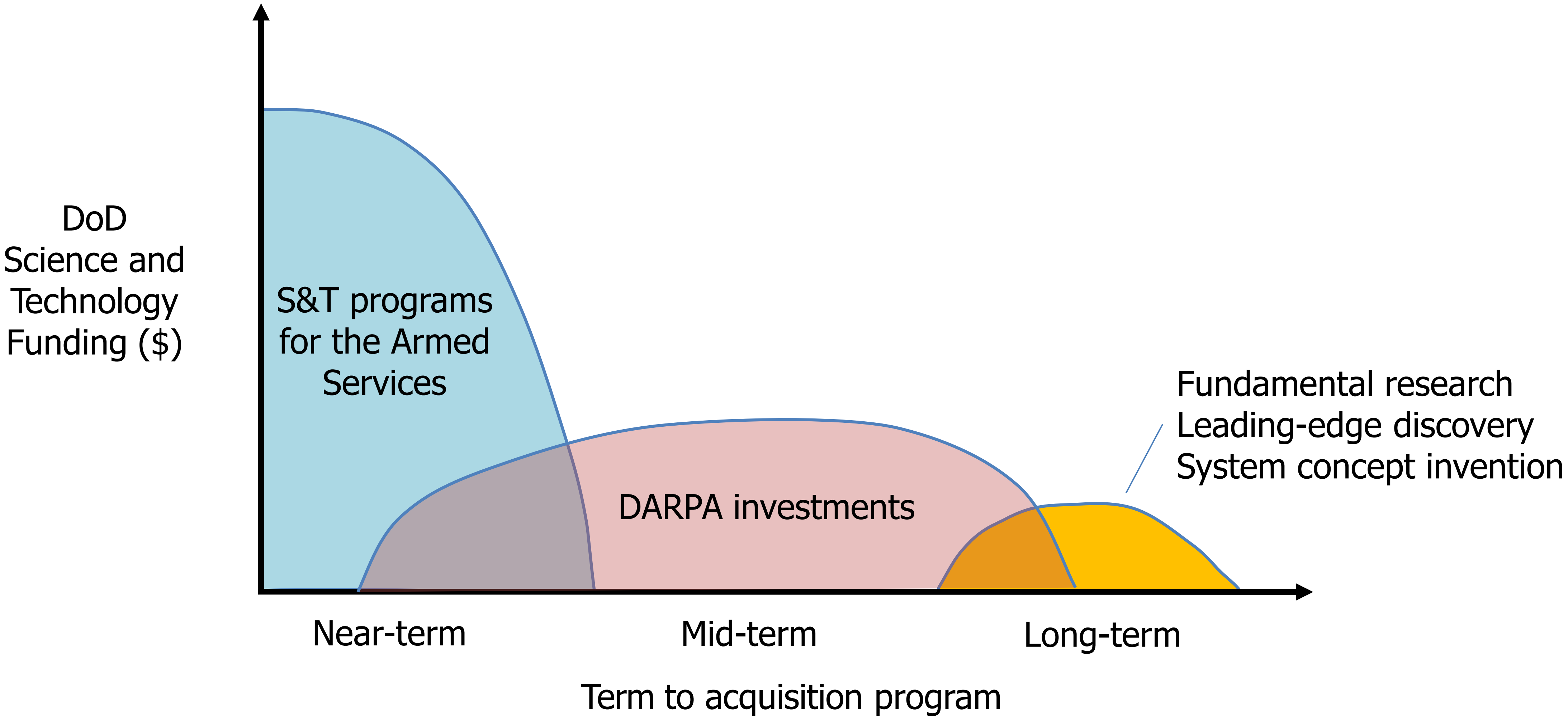
AI: Expectations and Realities

Dr. William Scherlis
Director, Information Innovation Office

March 2020

DARPA Achievements

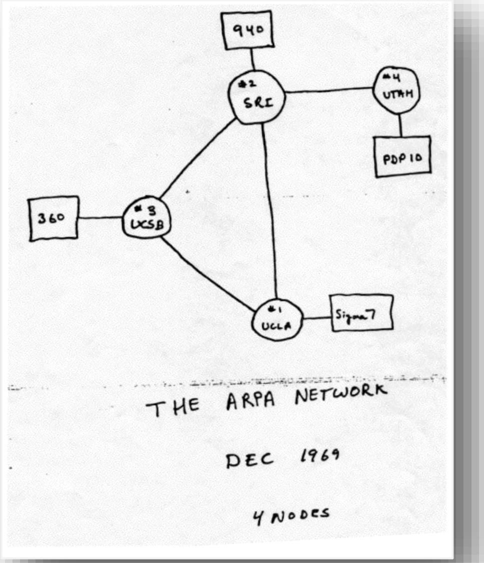
DARPA Role in Science and Technology



Information Innovation Office (I2O)

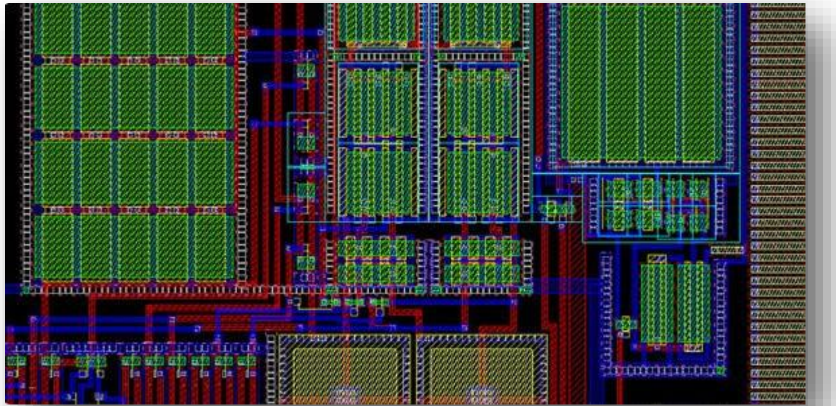
The first mouse

The internet



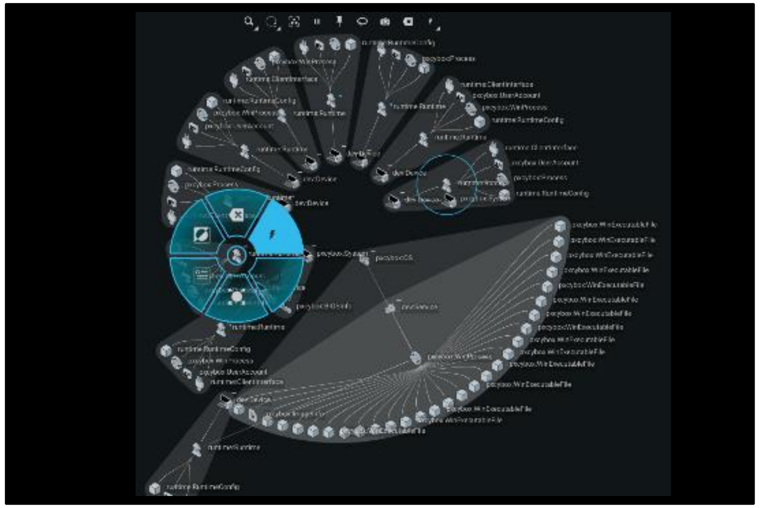
Foundations of artificial intelligence

Foundations for advanced ICs



Information Innovation Office (I2O)

Advantage in **cyber operations**



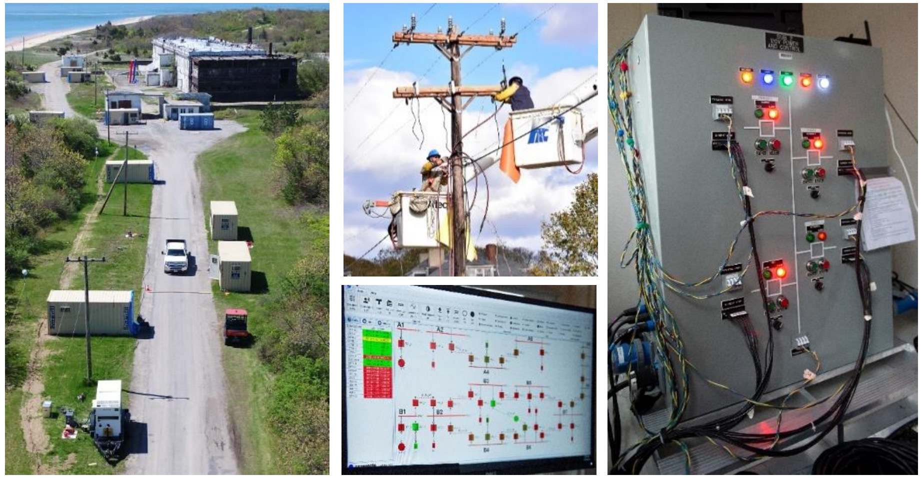
Artificial intelligence to the mission

Resilient, adaptable, and **secure systems**

Confidence in the **information domain**

Information Innovation Office (I2O)

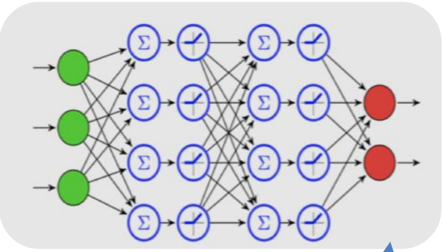
Advantage in cyber operations



RADICS

- Black start recovery of the power grid during a cyber attack

AI to the mission



Assured Autonomy

- *Explain*
- *Analyze/assure* - UUV

Function & safety cases

SMT solver

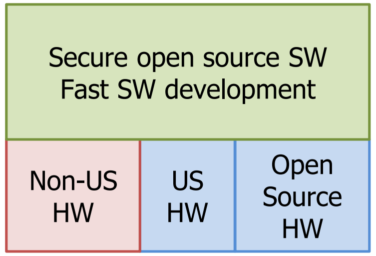
… (n == 5 + m) ∨ (p ∧ ¬q) …

SMT: Satisfiability modulo theories

Resilient, adaptable, and secure systems

OPS-5G

- *Open* – Hardware/software decoupling
- *Programmable* – Configure to the mission
- *Secure* – Trust and security



OPS-5G

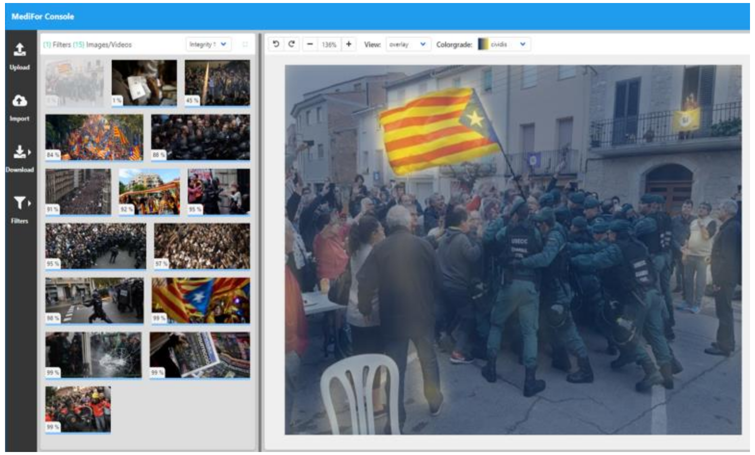
Confidence in the information domain

Media Forensics (MediFor)

- *Images/video* – Deep fakes

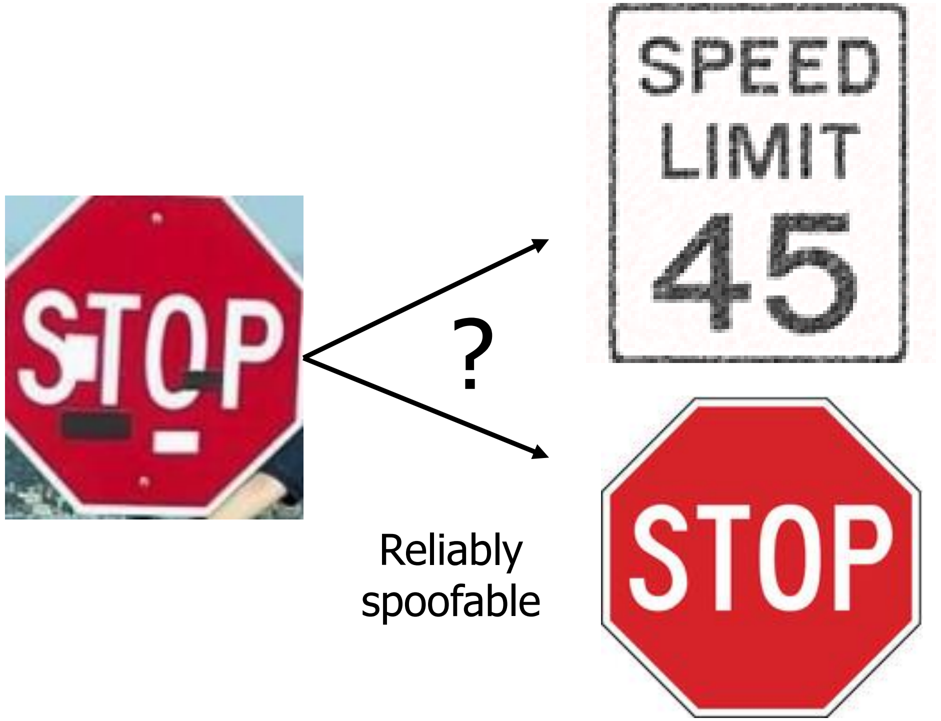
Semantic Forensics (SemaFor)

- *Multi-modal* – False narratives



Hard problems in AI

Machine learning: fragility, opacity, and dynamism



- How do we engineer systems to safely deliver AI to the mission?
- How do we harmonize domain models with AI techniques?
- How can humans best partner with AI-enhanced systems?
- What are successful models for continuous delivery, continuous integration and continuous verification for AI?
- What is next generation AI beyond symbolic and machine learning (waves 1 and 2 have shown their limits)?

The context of the I2O target operating environment

I2O Thrust Areas:

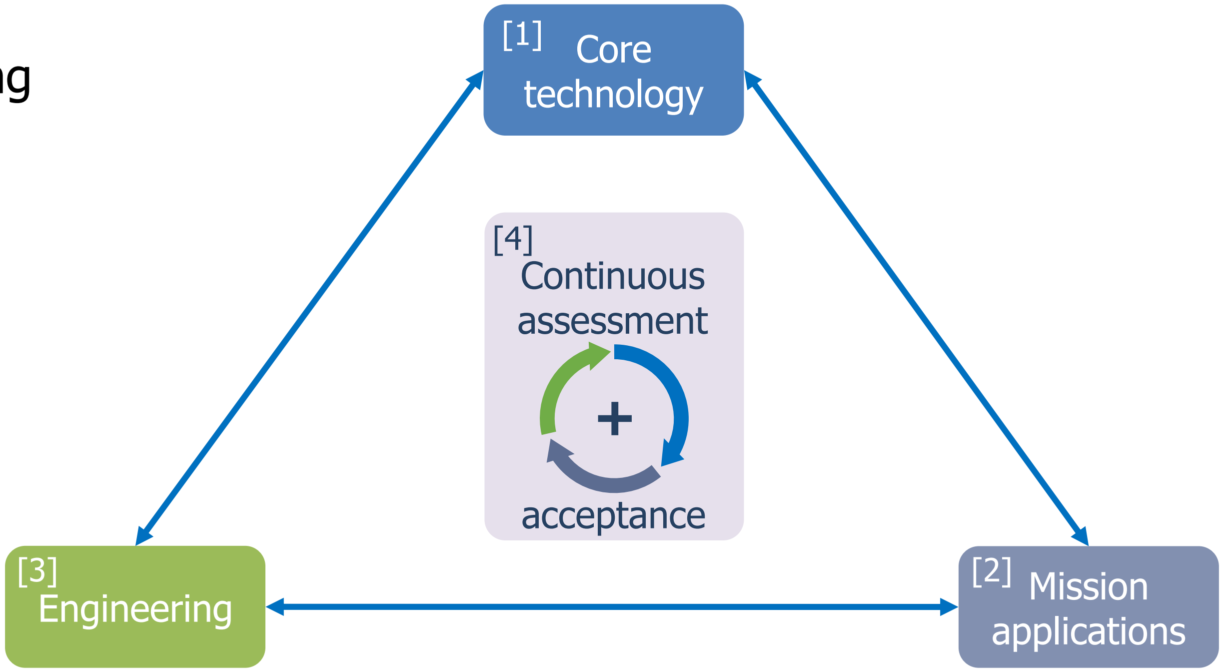
Artificial intelligence to the mission

Advantage in **cyber operations**

Resilient, adaptable, and **secure systems**

Confidence in the **information domain**

1. Today's adversaries (2+N) are capable and nimble
2. We must be able to continuously engineer at the margin – propelled by rapidly evolving threat and accelerating benefit from AI and computing
3. The technologies of I2O enable both defense and offense, with complex equities
4. With AI, human-machine partnering becomes more challenging



[1] Developing core AI

Context

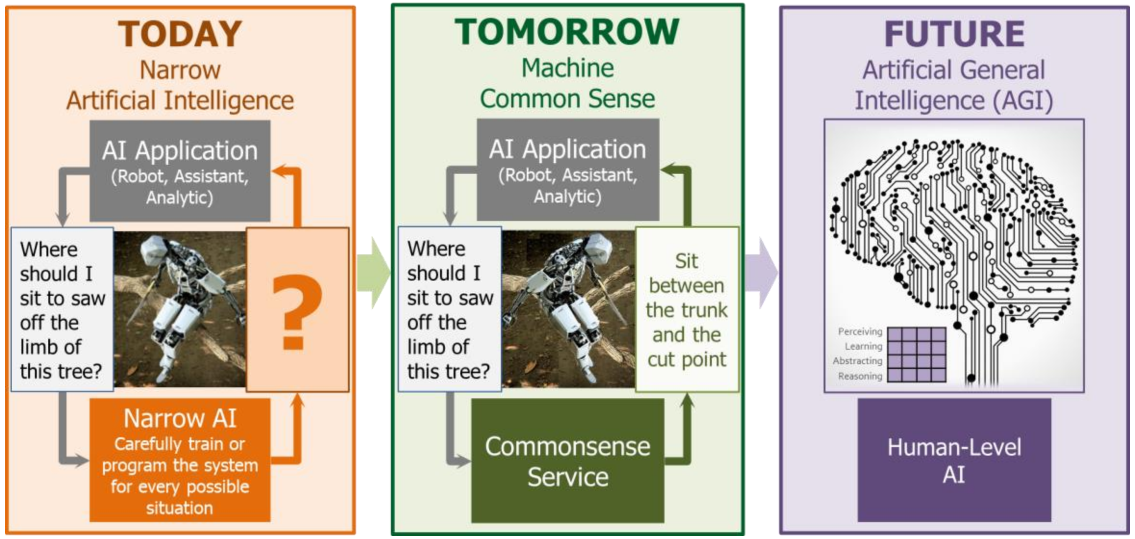
- Specialized cognitive building blocks: perception, reasoning, action

Approach

- Hybrids methods
  - Machine learning + game theory + optimization
  - Machine learning + explicit reasoning
- Infrastructure: Computing and data handling
- Looking ahead: Self adaptation – learning to learn

I2O Programs
- Communicating with Computers (CwC)
- Computers and Humans Exploring Software Security (CHESS)
- Explainable AI (XAI)
- Learning with Less Labeling (LwLL)
- **Machine Common Sense (MCS)**
- Synergistic Discovery and Design (SD2)

Machine Common Sense



Frame specialized AI using common sense reasoning

Enable AI applications to
- understand new situations,
- monitor the reasonableness of their actions
- transfer learning to new domains
- communicate more effectively with people

[2] Advancing mission applications of AI

Context

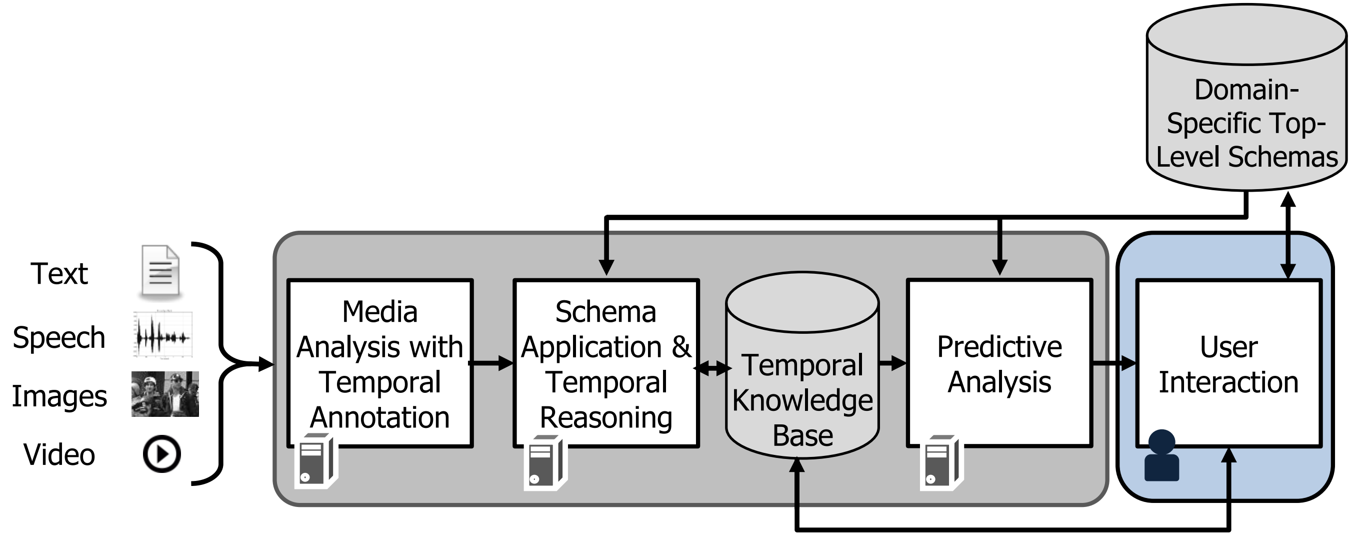
- Emerging AI-enabled mission concepts
- Adversaries are nimble and capable
- Human-AI partnering remains difficult
- Talent pool is a challenge

Approach

- Close partnering of operators and engineers
- Start with advisory AI

I2O Programs
- Active Interpretation of Disparate Alternatives (AIDA)
- Artificial Social Intelligence for Successful Teams (ASIST)
- Explainable AI (XAI)
- **Knowledge-directed AI Reasoning Over Schemas (KAIROS)**
- Media Forensics (MediFor)
- Semantic Forensics (SemaFor)

Knowledge-directed AI Reasoning Over Schemas (KAIROS)



Contextual and temporal reasoning

Create schema-based artificial intelligence capability to enable contextual and temporal reasoning about complex real-world events

[3] Engineering systems with embedded AI

Context

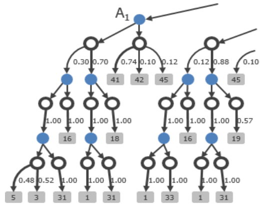
- Software and systems engineering are made more challenging with AI

Approach

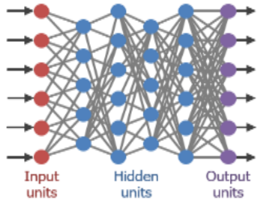
- Adapt key aspects of the engineering process
  - Integration frameworks, planning, and design
  - Process, tooling, and measurement
  - Assurance and evidence
  - Data, systems infrastructure, and configurations

I2O Programs
- Assured Autonomy (AA)
- **Explainable AI (XAI)**
- Symbiotic Design for Cyber Physical Systems (SDCPS)

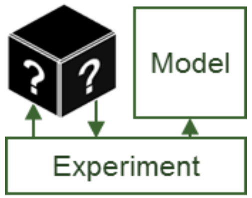
Explainable AI



Techniques to learn more structured, interpretable, causal models



Techniques to learn more explainable features



Techniques to infer an explainable model from any model as a black-box

Explain second wave AI

Enable human users to understand, trust, and effectively manage the emerging generation of AI partners

[4] Develop continuous evaluation and acceptance

Context

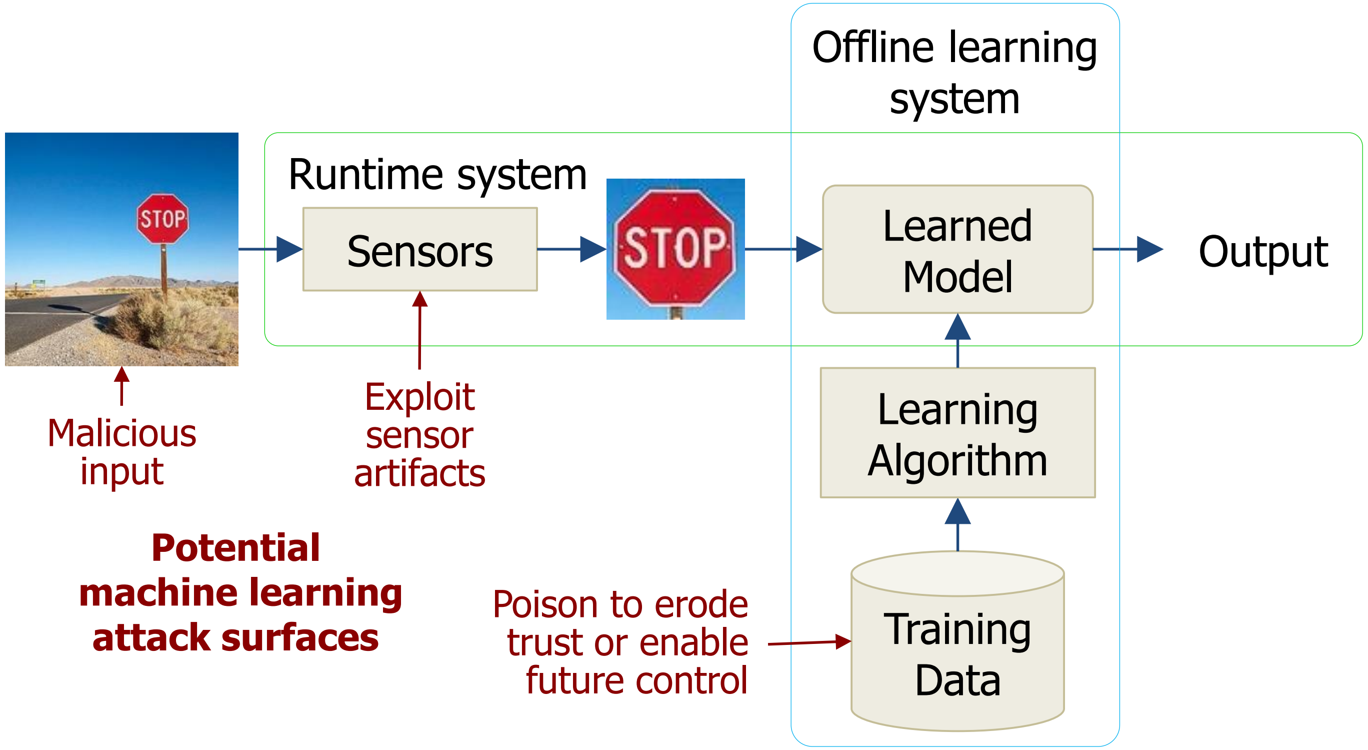
- Machine learning fragility, opacity, and dynamism
- Adversaries empowered in new ways, including attacking conventional systems
- Assurance influences all aspects of engineering and design, from the outset

Approach

- Integrate assurance planning
- Manage evidence to support confident accreditation decisions

I2O Programs
- Explainable AI (XAI)
- **Guaranteeing AI Robustness against Deception (GARD)**
- Media Forensics (MediFor)
- Semantic Forensics (SemaFor)

Guaranteeing AI Robustness against Deception (GARD)



Design robust and resilient AI models

Enable machine learning systems to be robust against adversary deception www.darpa.mil