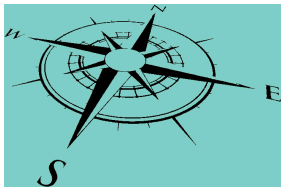
AI: Expectations and Realities

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)?

Contents



Information Innovation Office (I2O)

Description:

* Advantage in cyber operations

- Artificial intelligence to the mission
- Resilient, adaptable, and secure systems
- Confidence in the information domain

Stakeholder(s):

DARPA :
DARPA Achievements:

- *MATERIAL SCIENCE *GPS*
- *NIGHT VISION*
- *PREDATOR AUTONOMOUS VEHICLE*
- *NEUROPROSTHETIC LIMBS*
- *STEALTH FIGHTER*
- *ADVANCED SEMICONDUCTORS*
- *SATURN V*
- *AUTONOMOUS VEHICLES*
- *PERSONALIZED ASSISTANT THAT LEARNS*

Dr. William Scherlis :
Director, Information Innovation Office

I2O Programs

Communicating with Computers (CwC) Program

Computers and Humans Exploring Software Security (CHESS) Program

Explainable AI (XAI) Program

Learning with Less Labeling (LwLL) Program

Machine Common Sense (MCS) Program

Synergistic Discovery and Design (SD2) Program

Vision

3G AI

Mission

To solve hard problems in AI

Values

Resilience: Resilient, adaptable, and secure systems

Adaptability

Openness: Open – Hardware/software decoupling

Programmability: Programmable – Configure to the mission

Security: Secure – Trust and security

1. AI Core

Develop core AI

Stakeholder(s)

Machine Common Sense (MCS) Program

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

Frame specialized AI using common sense reasoning.

1.1. Situational Understanding

Enable AI applications to understand new situations

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

1.2. Reasonableness

Enable AI applications to monitor the reasonableness of their actions

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

1.3. Learning Transfer

Enable AI applications to transfer learning to new domains

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

1.4. Communication

Enable AI applications to communicate more effectively with people

Stakeholder(s):

People

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

2. Applications

Advance mission applications of AI

Stakeholder(s)

Knowledge-directed AI Reasoning Over Schemas (KAIROS) Program

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

2.1. Reasoning

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

Text | Speech | Images | Video > Media Analysis with Temporal Annotation > Schema Application & Temporal Reasoning > Temporal Knowledge Base > Predictive Analysis > User Interaction

Performance Indicators

2.1.1a Text

Description	Type	Start Date	End Date
Capture text as input	Target		
	Actual		

Relationships

Speech -

Images -

Video -

Feeds into Analysis & Annotation -

2.1.1b Speech

Description	Type	Start Date	End Date
Capture speech as input	Target		
	Actual		

Relationships

Text -

Images -

Video -

Feeds into Analysis & Annotation -

2.1.1c Images

Description	Type	Start Date	End Date
Capture images as input	Target		
	Actual		

Relationships

Text -

Speech -

Video -

Feeds into Analysis & Annotation -

2.1.1d Video

Description	Type	Start Date	End Date
Capture video as input	Target		
	Actual		

Relationships

Text -

Speech -

Images -

Feeds into Analysis & Annotation -

2.1.2 Analysis & Annotation

Description	Type	Start Date	End Date
Media Analysis with Temporal Annotation	Target		
	Actual		

Relationships

Feeds into Applications & Reasoning -

2.1.3 Applications & Reasoning

Description	Type	Start Date	End Date
Schema Application & Temporal Reasoning	Target		
	Actual		

Relationships

Feeds into Knowledge Base -

2.1.5 Knowledge Base

Description	Type	Start Date	End Date
Temporal Knowledge Base	Target		
	Actual		

Relationships

Supports Prediction -

Supports Interaction -

2.1.6 Prediction

Description	Type	Start Date	End Date
Predictive Analysis	Target		
	Actual		

Relationships

Facilitates Interaction -

2.1.7 Interaction

Description	Type	Start Date	End Date
User Interaction	Target		
	Actual		

Relationships

Feeds Back into the Knowledge Base -

Feeds into the Schemas -

2.1.8 Schemas

Description	Type	Start Date	End Date
Domain-Specific Top-Level Schemas	Target		
	Actual		

Relationships

Support Prediction -

Support Applications & Reasoning -

3. Systems Engineering

Engineer systems with embedded AI

Stakeholder(s)

Explainable AI (XAI) Program

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

3.1. Understanding, Trust & Management

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

Explain second wave AI

Stakeholder(s):

Humans **AI Partners**

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

3.1.1. Causal Models

Learn more structured, interpretable, causal models

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

3.1.2. Features

Learn more explainable features

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

3.1.3. Black-Box Models

Infer an explainable model from any model as a black-box

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

4. Evaluation & Acceptance

Develop continuous evaluation and acceptance

Stakeholder(s)

Guaranteeing AI Robustness against Deception (GARD) Program

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

4.1. Deception

Enable machine learning systems to be robust against adversary deception

Design robust and resilient AI models

Performance Indicators

Description	Type	Start Date	End Date
	Target		
	Actual		

Administrative Information

Start Date:
End Date:

Publication Date: 2020-03-30
Source: https://download.1105media.com/Custom/Workshops/2020/AI/Bill_Scherlis.pdf

Submitter:

Given Name: Owen
Surname: Ambur
Email: Owen.Ambur@verizon.net
Phone:

_00efbeea-72ad-11ea-9d53-b7df97babdf6