



S Y S T E M S
E N G I N E E R I N G
R E S E A R C H C E N T E R

AIRC

ACQUISITION INNOVATION
RESEARCH CENTER

SE IN THE ERA OF HUMAN- MACHINE TEAMING ROADMAP FOR AI AND SE

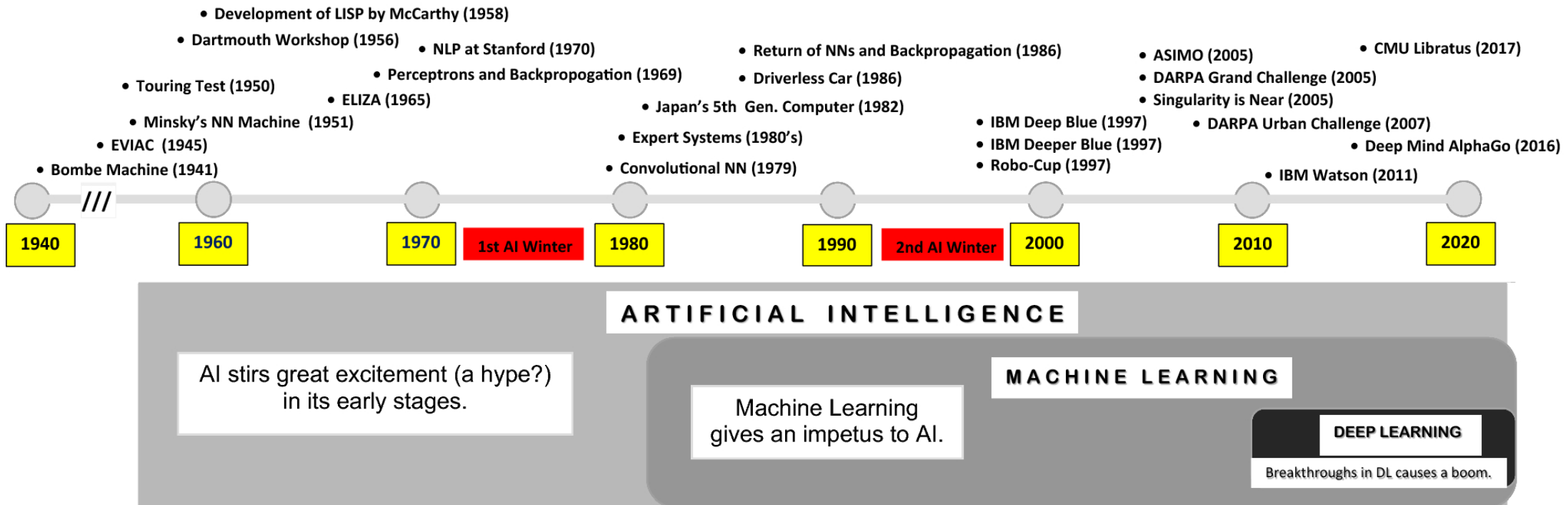
Tom McDermott, Systems Engineering Research Center



AGENDA

1. SE4AI and AI4SE and the SERC Research Roadmap
2. Systems Engineering and AI
3. Human-Machine Teaming

AI/ML IS NOT NEW TECHNOLOGY



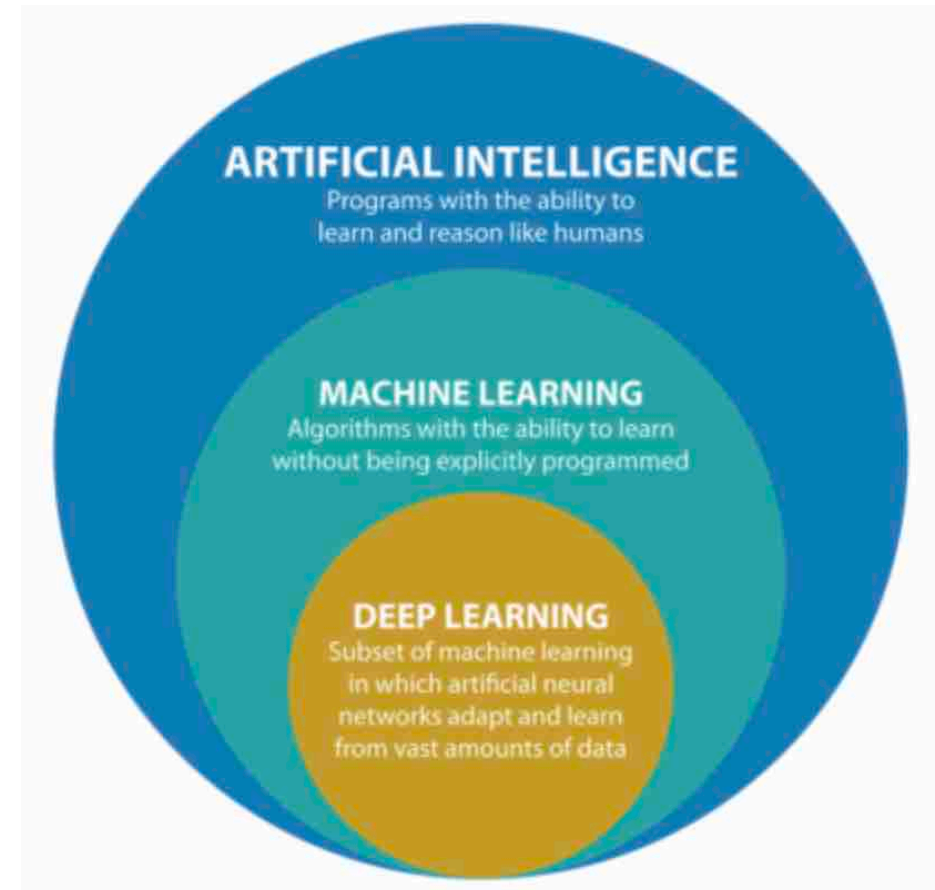
<https://link.springer.com/article/10.1007/s44163-021-00009-x/figures/1>

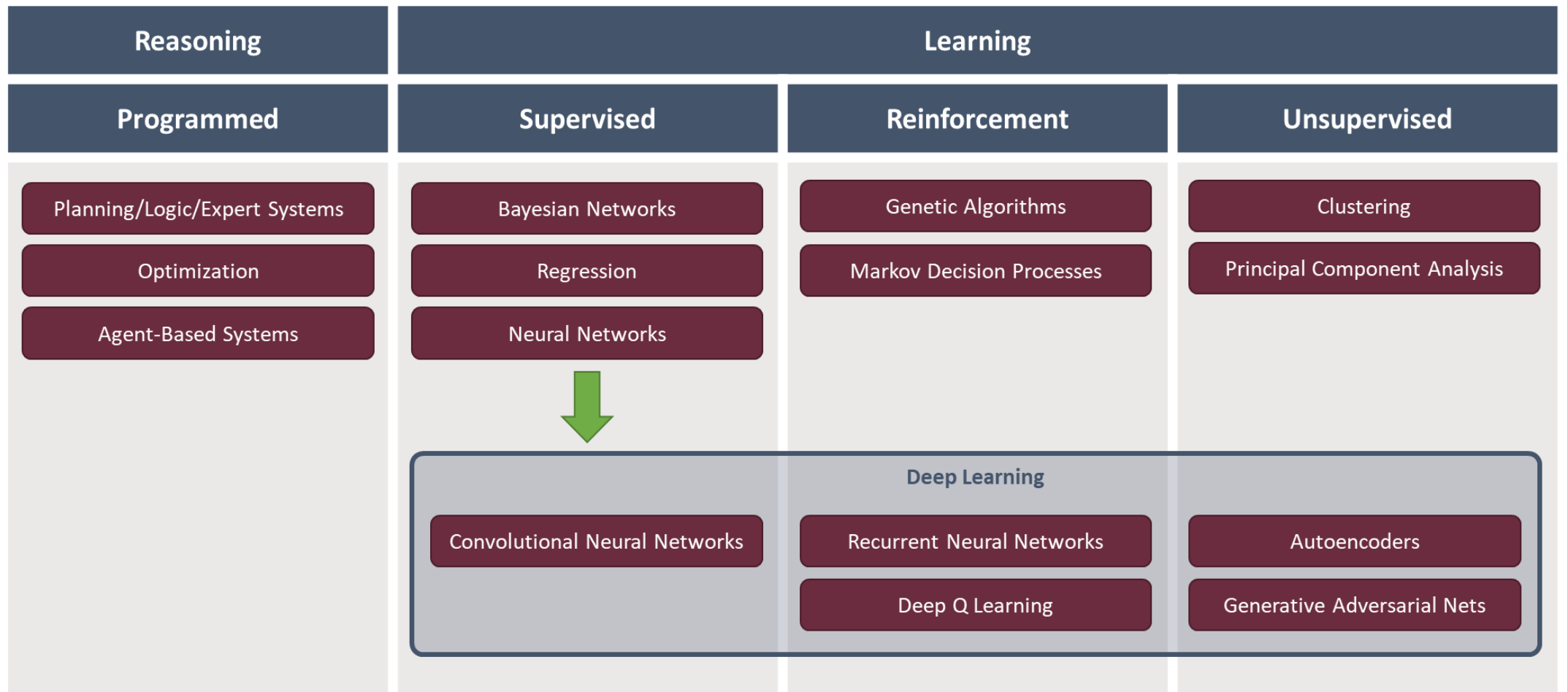
derives from
cognitive science

derives from **statistical
data modeling**

	AI	ML
1	AI allows a machine to simulate human intelligence to solve problems	ML allows a machine to learn autonomously from previous data
2	The goal is to develop an intelligent system that can perform complex tasks	The goal is to build machines that can learn from data to increase the accuracy of the output
3	AI uses technologies in a system so that it mimics human decision-making	ML uses self-learning algorithms to produce predictive models
4	AI works with all types of data: structured, semi-structured, and unstructured	ML can only use structured and semi-structured data
5	AI systems use logic and decision trees to learn, reason, and self-correct	ML systems rely on statistical models to learn and can self-correct when provided with new data

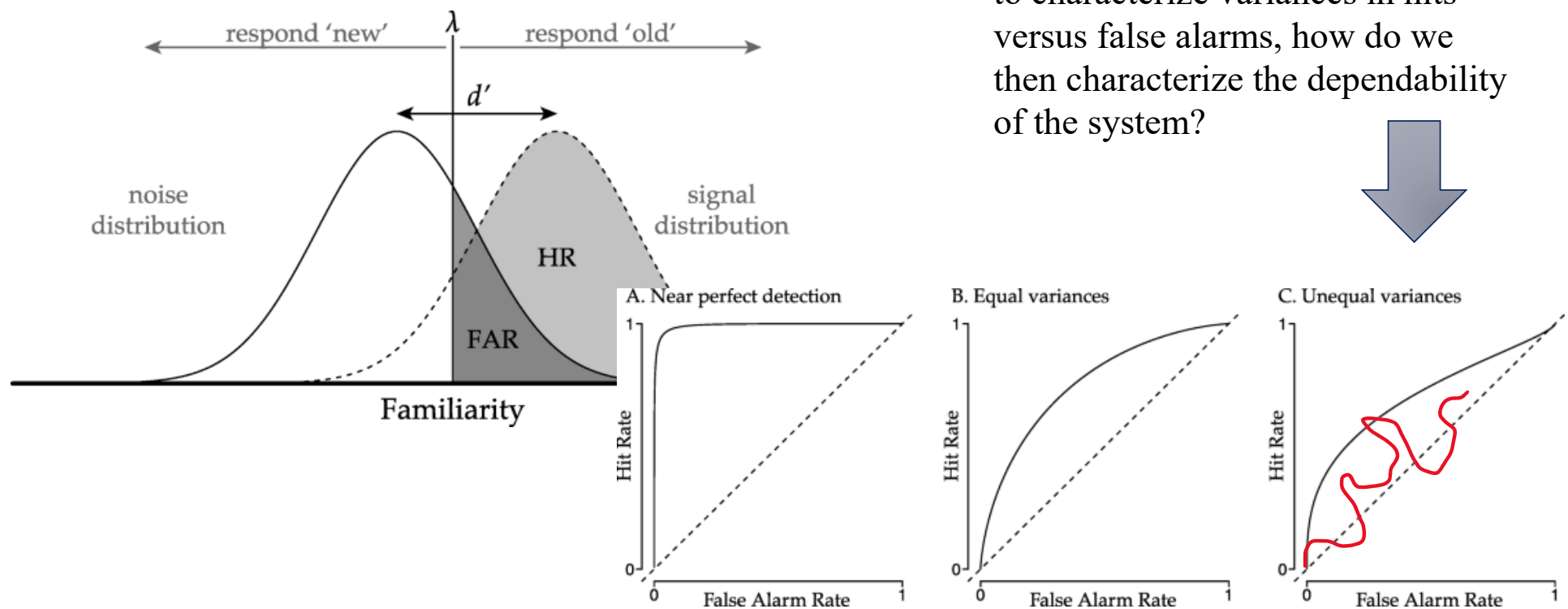
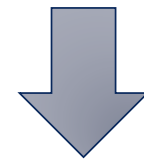
AI AND ML ARE NOT THE SAME THING





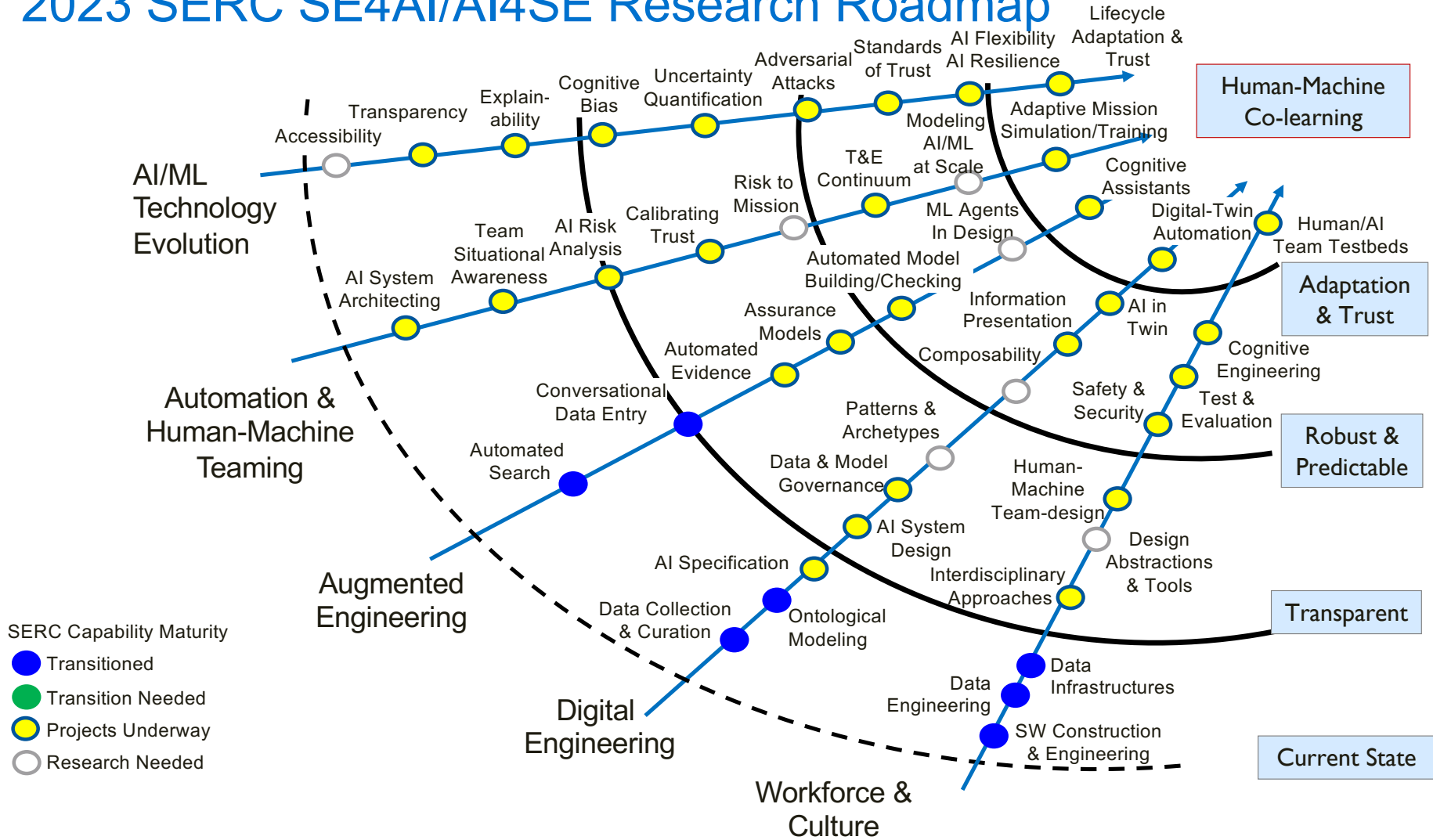
CHARACTERIZING ERROR

If we have no deterministic model to characterize variances in hits versus false alarms, how do we then characterize the dependability of the system?



Figures from: Selker, R., van den Bergh, D., Criss, A.H. *et al.* Parsimonious estimation of signal detection models from confidence ratings. *Behav Res* **51**, 1953–1967 (2019).

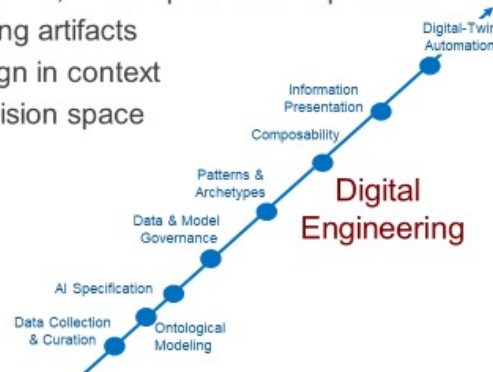
2023 SERC SE4AI/AI4SE Research Roadmap



DIGITAL ENGINEERING TRANSFORMATION

AI Enabled Digital Engineering

- **Data Collection and Curation** - data collection, management, curation and governance
- **Ontological Modeling** – schematic representation to semantic representation
- **Specification** – what will be allocated to the machine, in both product and process
- **Patterns and Archetypes** – learning from modeling artifacts
- **Composability** – training and evaluating for design in context
- **Information Presentation** – representing the decision space for human understanding and learning
- **Digital Twin Automation** – real-time continuous learning from real system and shadow simulations



Convergence of Data Science and Systems Engineering Disciplines

Models become central to defining complex systems of systems

Results in Product plus **Virtual Twins** of Product

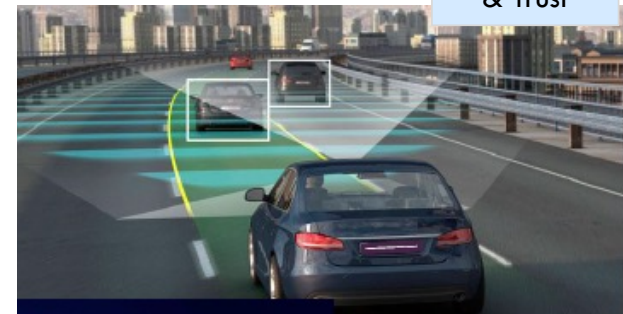
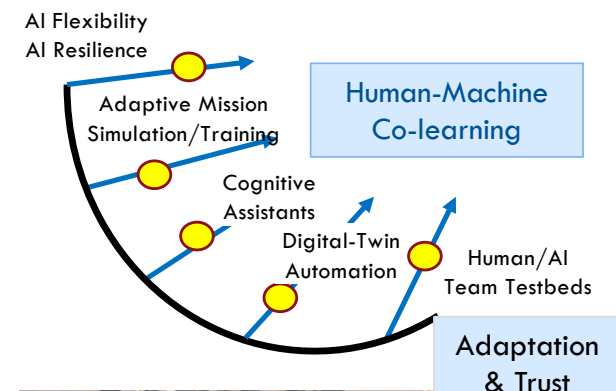
Human-Machine interfaces and **Visualization** of complex interrelationships

HUMAN-MACHINE CO-LEARNING

Adaptive Cyber-Physical-Human Systems – digital twins: modeling of cyber-physical systems as influenced by humans, in testbeds...

Adaptive Mission Simulation/Training – Simulation and training that supports non-static objectives (pick-up games)

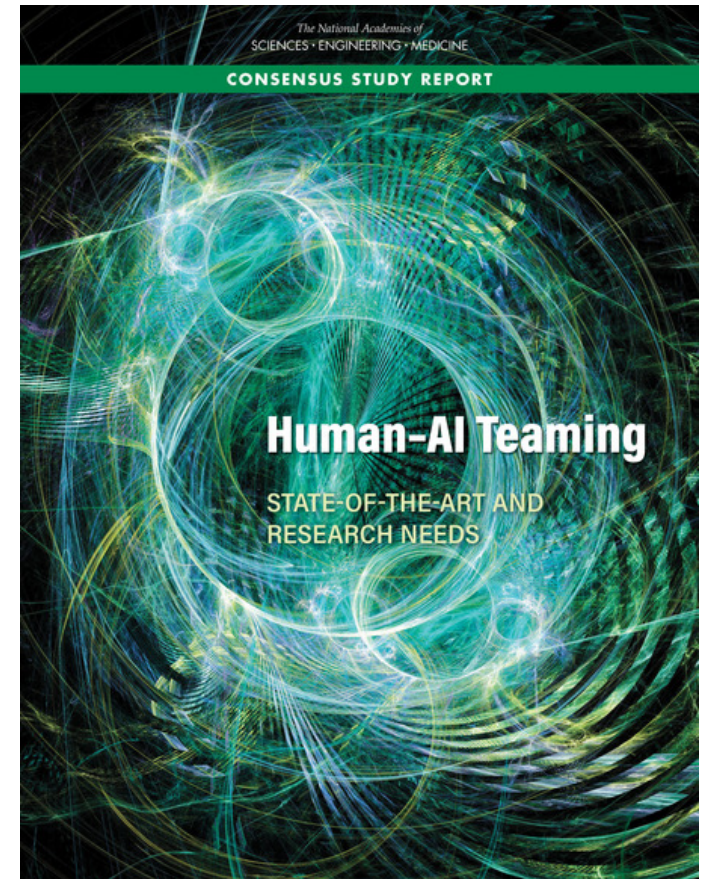
AI Flexibility & Resilience – AI systems that self-adapt to changing operational boundaries while maintaining rigorous safety and security and policy constraints



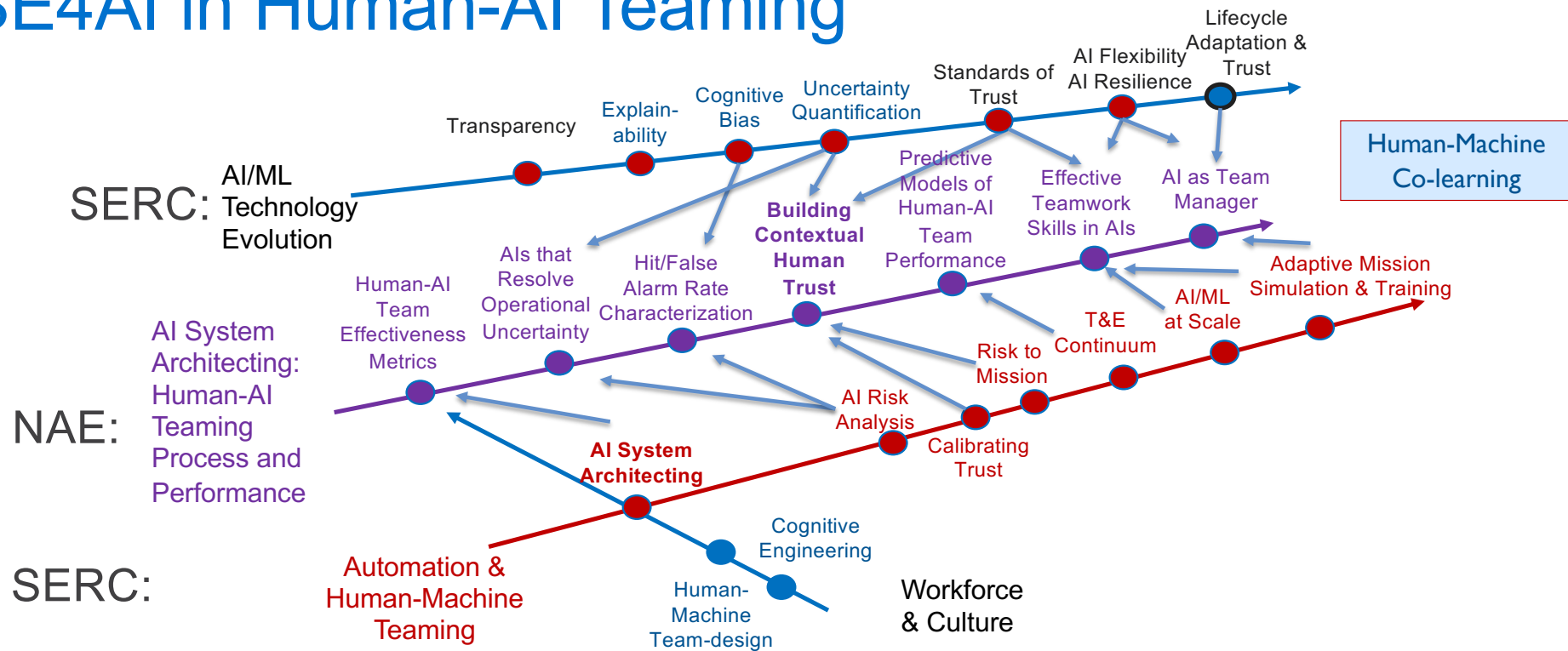
SE/HSI Objectives

Significant value in considering the human and AI as a team

- Long-term, distributed, and agile human-AI teams through improved team assembly, goal alignment, communication, coordination, social intelligence, and the development of a new human-AI language – **AI System Architecting**
- Methods for improving human situational awareness of AI systems
- Improved AI system transparency and explainability
- **Interaction mechanisms and strategies within the human-AI team**
- Advance understanding of how broader sociotechnical factors affect trust in human-AI teams
- Better understand the interdependencies between human and AI decision-making biases
- What, when, why, and how to best train human-AI teams
- **Advances in HSI processes and measures**



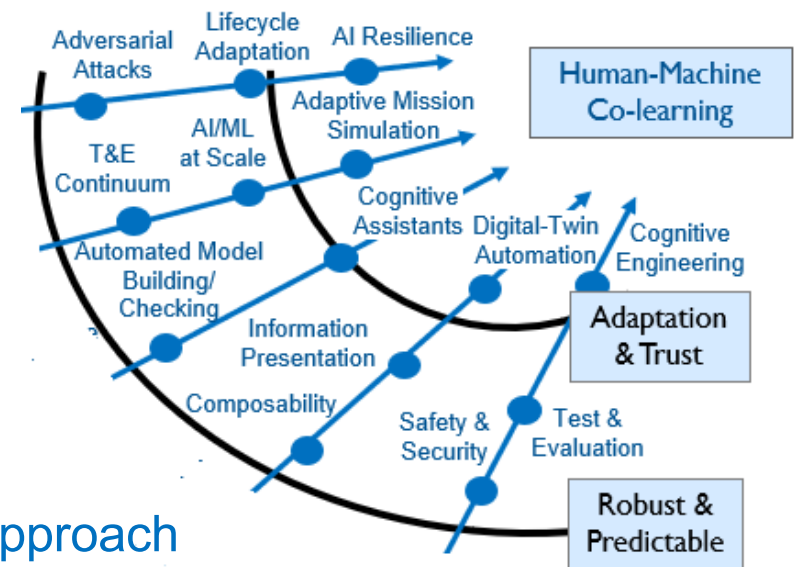
SE4AI in Human-AI Teaming



- Long-term, distributed, and agile human-AI teams through improved team assembly, goal alignment, communication, coordination, social intelligence, and the development of a new human-AI language – **AI System Architecting**
- What, when, why, and how to best train human-AI teams
- Advances in HSI processes and measures

Challenges for Test & Evaluation of AI

- Testing & Evaluation is a **continuum**
 - Information accumulates over time across varying operating envelopes
 - does not end until the system retires
- All AI areas need **testbeds**
- Operational relevance is essential
- Data Management is foundational
- AI systems require a **probabilistic risk-based approach**
- Previous test metrics apply, but may have different interpretations
 - Task & mission level performance, course of action, non-functional requirements
- An expanded definition of **external context** is necessary
- The T&E workforce and culture must evolve



Freeman, L. (2020), Test and Evaluation for Artificial Intelligence. INSIGHT, 23: 27-30.

Holistic view of the system of systems

Measurement of “ilities” (e.g., flexibility, resilience, trust)

Architecting / Human-system integration

Product platforms / evolvability of systems of systems

Lifecycle risk analysis

Linking “Design for X” “T&E” and lifecycle value.

Understanding human behavior as part of the system

Emergent system behavior

Building user Trust by understanding the Human AI system

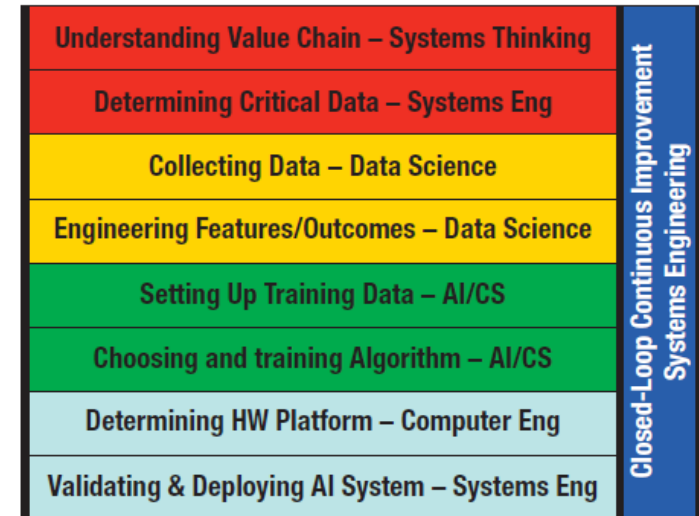
Architecting AI Systems for long-term trust: Linking task & function allocation, test and risk analysis and need for systems testbeds

T&E as a Continuum: what to test and how to interpret for AI Systems of varying complexity and embeddedness

AI Resilience: Strategies to mitigate disruptions / ensure acceptable behaviors and recoveries when failures occur

Workforce and Culture

- Digital Engineering Competencies
- Integrating AI/ML experts with Domain experts, all disciplines
- Evolving tools to align with design and disciplinary abstractions =>
- **Human Systems Engineering: no longer a specialty discipline**
- Threat models, safety, security, resilience, and other 'ilities
- Evolving test and evaluation competency
- Training the Users to appropriately interact with AI's



Wade, J., Buenfil, J. and Collopy, P. (2020), A Systems Engineering Approach for Artificial Intelligence: Inspired by the VLSI Revolution of Mead & Conway. INSIGHT, 23: 41-47.



SERC 5TH ANNUAL AI4SE & SE4AI WORKSHOP



2023 SUMMARY REPORT



The conference theme, “Safer AI-Enabled Complex Systems: Responsible Deployment of AI through Systems Engineering,” aims to foster discussions and insights on how systems engineering can support the development of robust and ethical AI systems, and how AI tools can in turn transform the practice of systems engineering.

Abstract submissions through 17 June 2024

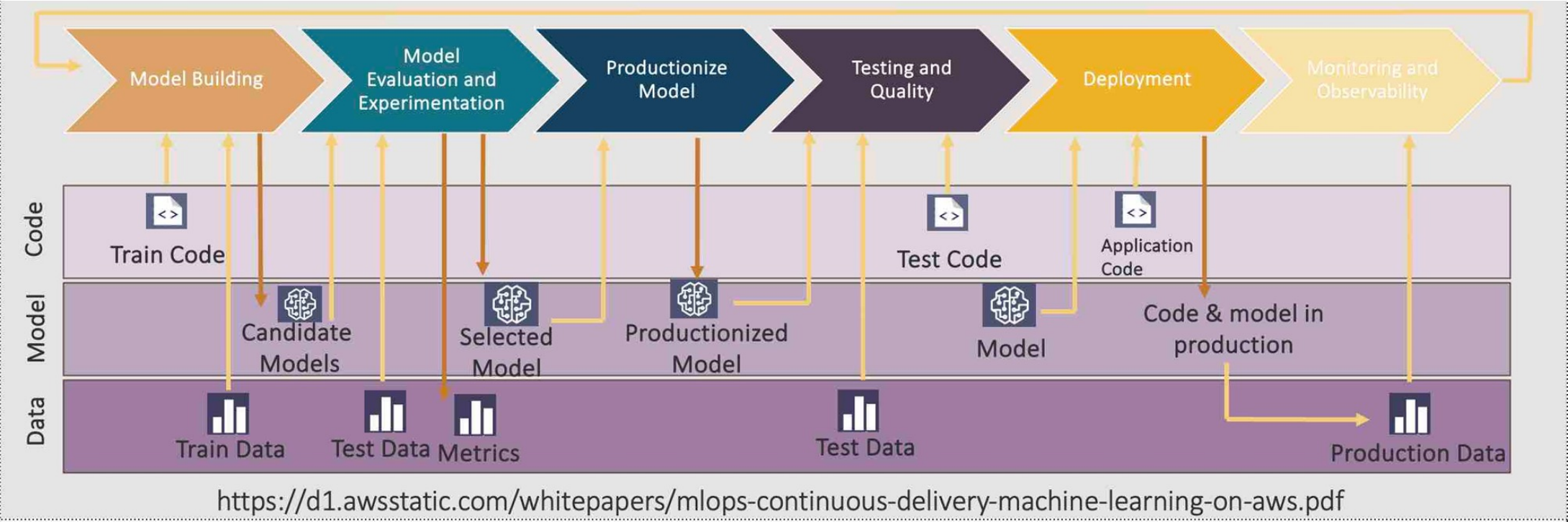
<https://sercuarc.org/event/ai4se-se4ai-workshop-2024/#dates>



AGENDA

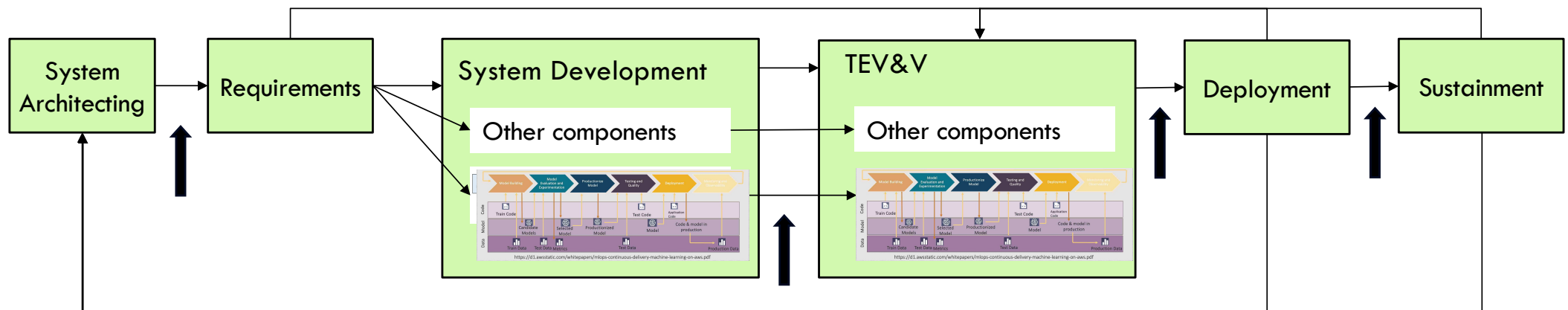
1. SE4AI and AI4SE and the SERC Research Roadmap
2. Systems Engineering and AI
3. Human-Machine Teaming

Typical representation of AI/ML pipeline:



... but this is still focused on the AI model as the system.

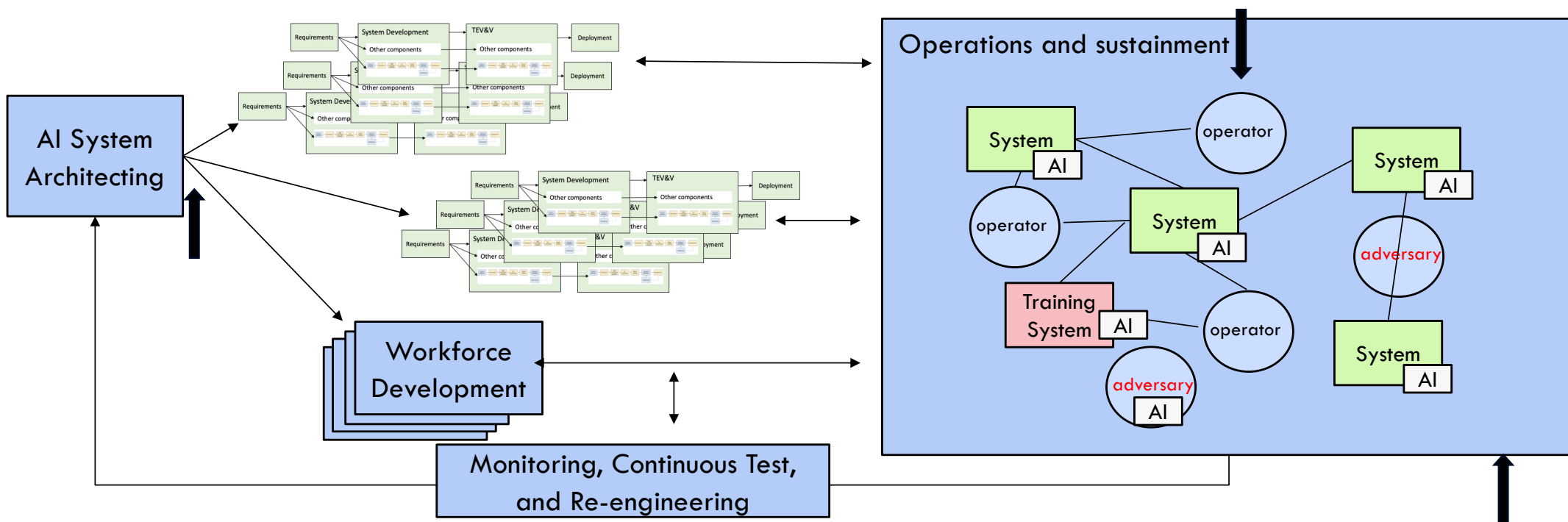
For Systems Engineers, AI is part of a "system"



Emphasizes tradeoffs in performance and risk

Recognizes that system might need to work in unplanned ways over its lifecycle and that behavior (and failures) must be acceptable

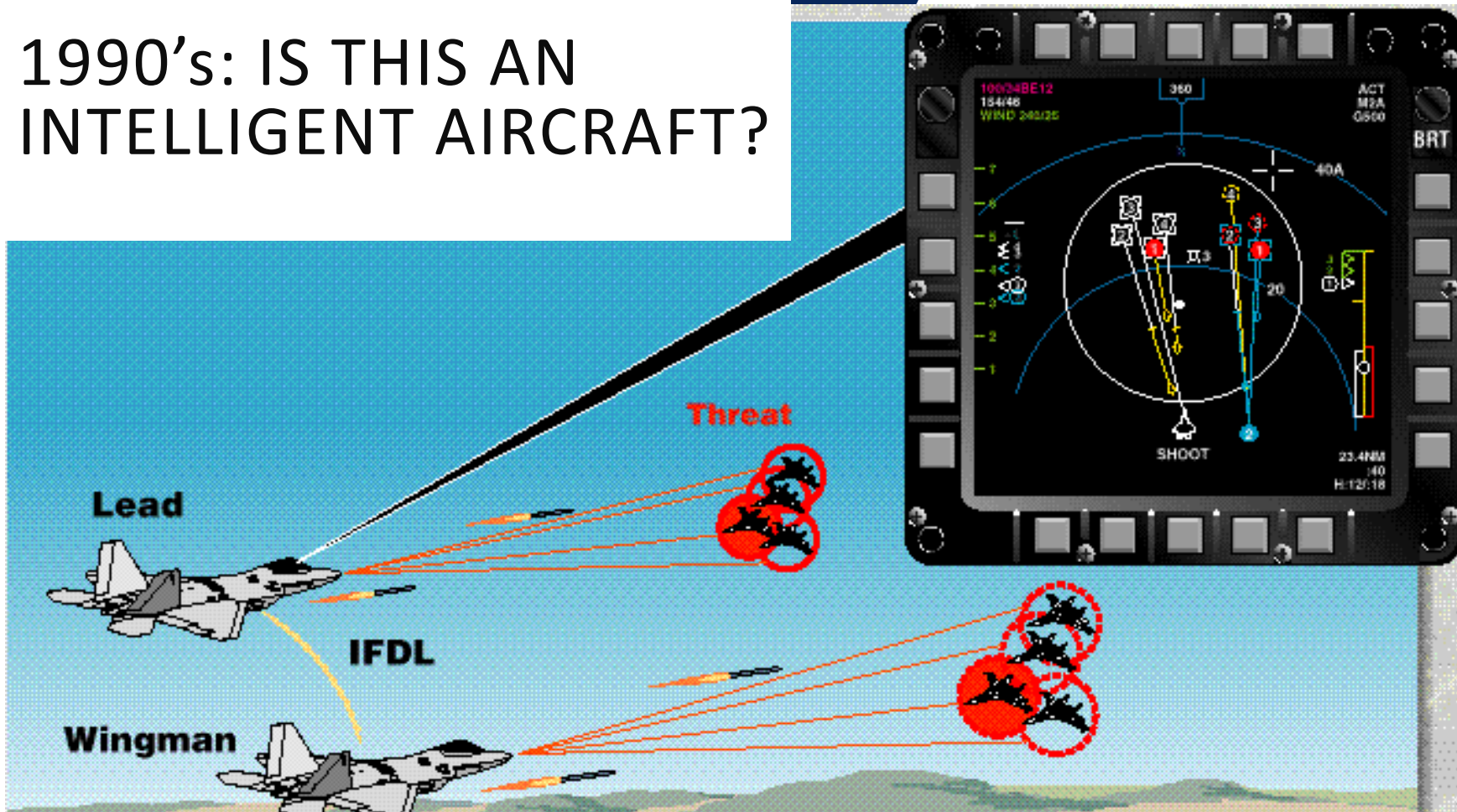
The real world operates in a socio-technical systems environment, involving complex interactions among humans and systems that were not always intended to work together in a constantly changing environment.



Everything on the previous slides... and extent to which operators use and trust new technology, how risks and functions are allocated to different parts of the overall systems, how changing environment is monitored, and network is updated accordingly

1. SE4AI and AI4SE and the SERC Research Roadmap
2. Systems Engineering and AI
3. **Human-Machine Teaming**
 1. Building user Trust by understanding the Human AI system
 2. Architecting AI Systems for long-term trust: Linking task & function allocation, test and risk analysis and need for systems testbeds
 3. T&E as a Continuum: what to test and how to interpret for AI Systems of varying complexity and embeddedness
 4. AI Resilience: Strategies to mitigate disruptions / ensure acceptable behaviors and recoveries when failures occur

1990's: IS THIS AN INTELLIGENT AIRCRAFT?



The sensor fusion loop detects threat aircraft, tracks location and movement, identifies the type, calculates an optimal engagement, even tells the pilot when to shoot. The pilot must initiate the shot. This all happens beyond the visual range of the pilot. How does the pilot trust the information provided by the sensor fusion in this critical situation?

Developer

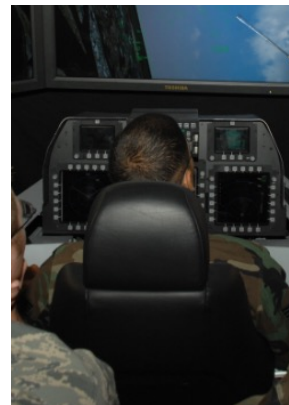


[1]

Accuracy:

If you're a computer scientist you want to see the math of this specific algorithm or at least a visualization of the prediction.

Domain Expert



[2]

Agrees with me:

If you're a pilot flying in an engagement using your display image, you might want to see the system agree with you often enough.

End User



[3]

Trusted 3rd Party:

If you're an operational evaluator, you might want to certify it's safety...and for commanders, not have created any international incidents!

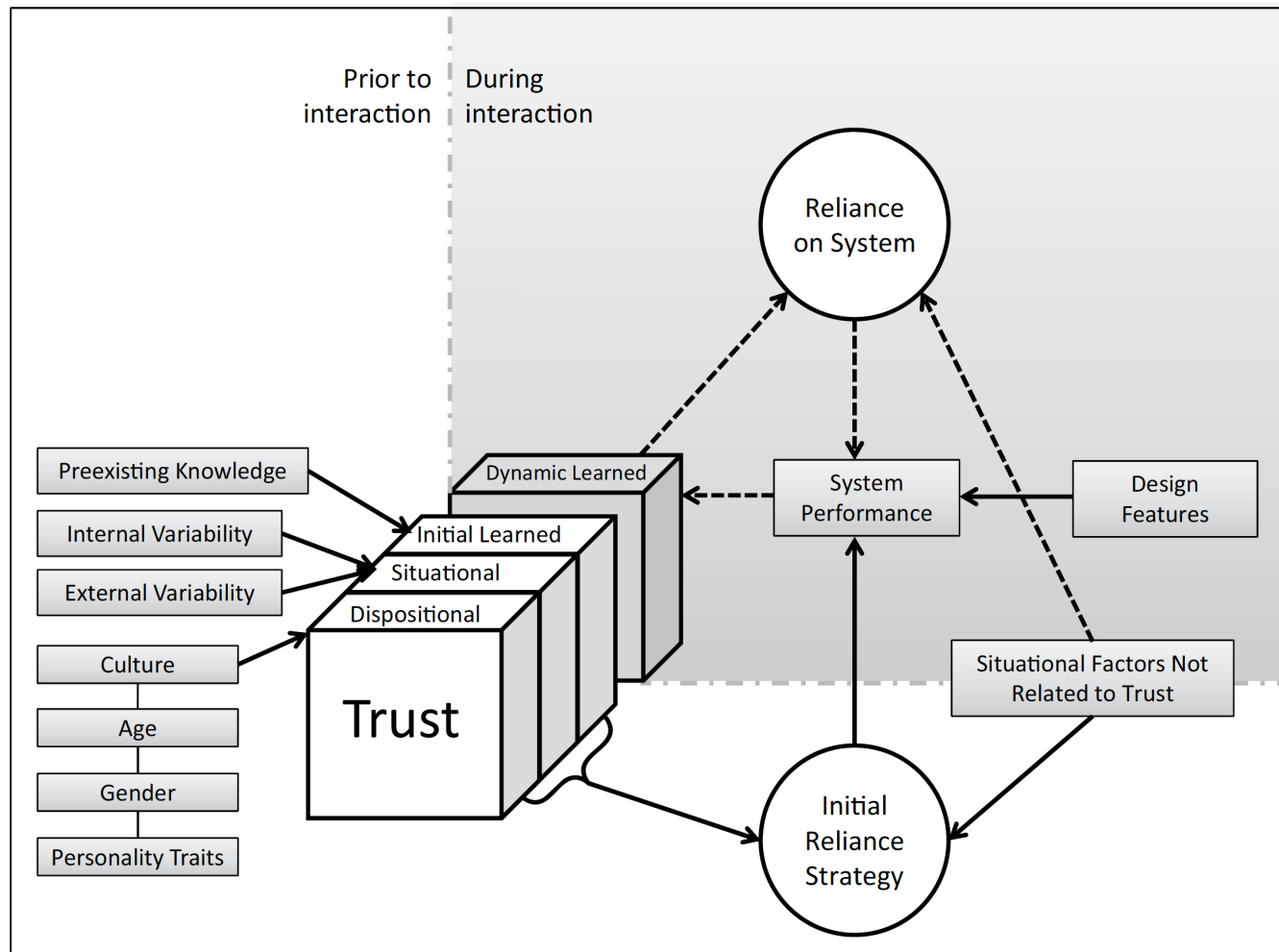
Simulators and test systems allow the user to evaluate the system behaviors in a larger context, to make a judgement about the final decision action.

CONTEXTUAL TRUST



ENGINEERED TRUST

Transparency in the underlying algorithms and behaviors created by engaging the user in the development process and matured in critical reviews.



Hoff and Bashir's model of factors influencing trust in automation.

This focuses on one human and one autonomous agent. What happens when there are multiple humans and multiple agents working in different phases of the system?

Replacing/augmenting existing task



[4]

Developer:	Domain Expert:	End User:
Inspect algorithm	Compare to what I would do	Reputable source (logo/medallion)

Replacing/augmenting existing task



Solving new system level problem



[5]

What should the answer look like?

1. SE4AI and AI4SE and the SERC Research Roadmap
2. Systems Engineering and AI
3. **Human-Machine Teaming**
 1. Building user Trust by understanding the Human AI system
 2. **Architecting AI Systems for long-term trust: Linking task & function allocation, test and risk analysis and need for systems testbeds**
 3. T&E as a Continuum: what to test and how to interpret for AI Systems of varying complexity and embeddedness
 4. AI Resilience : Strategies to mitigate disruptions / ensure acceptable behaviors and recoveries when failures occur



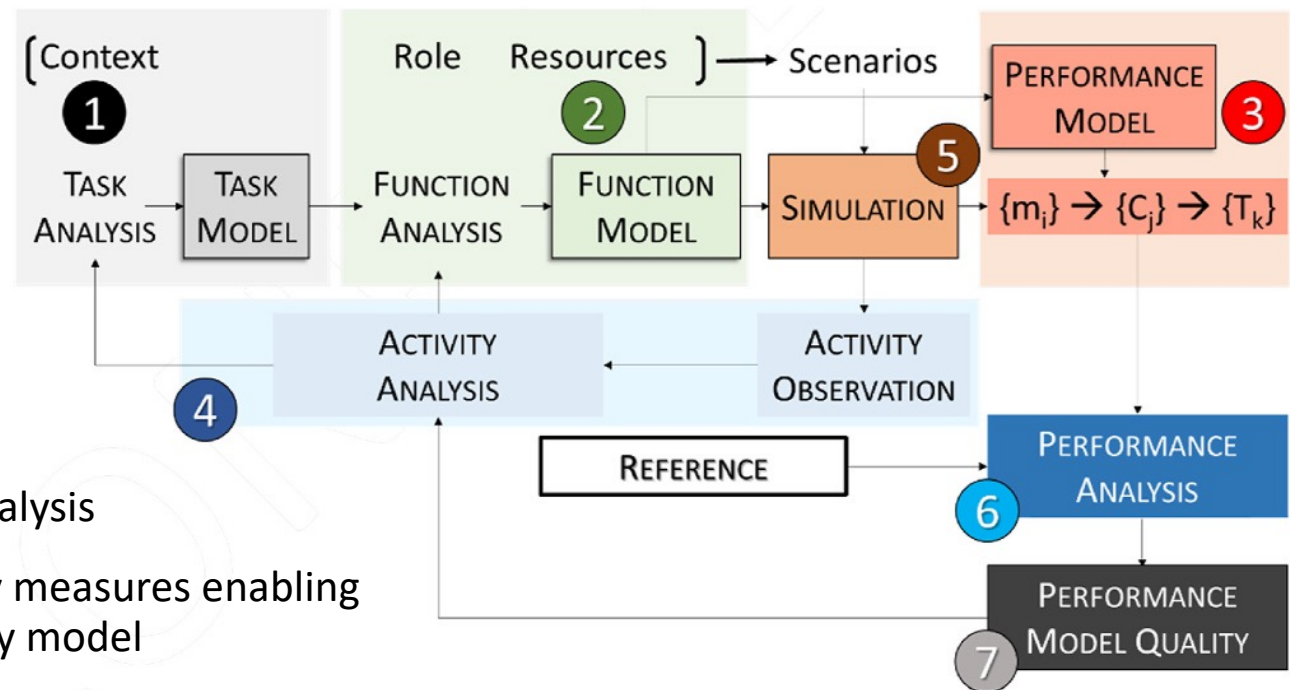
What are the Human Tasks?
Machine Tasks?

What are the Human Functions?
Machine Functions?

What are the flows of Information
between these?

What is the expected and measured
performance?

1. Task analysis enabling the building of a task model
2. Function analysis enabling the construction of a functional model
3. Performance analysis based on a performance model
4. Activity analysis to elaborate the function model
5. Human-in-the-loop simulations can be performed, and human and machine activity can be observed, enabling an activity analysis
6. Leading to a system performance analysis
7. Evaluated using performance quality measures enabling the building of a performance quality model



Boy GA, Masson D, Durnerin É, Morel C. PRODEC for human systems integration of increasingly autonomous systems. Systems Engineering. 2024;1-22

Countermine Operational Scenario



- Strait of Hormuz, suspected minefield threatens open Sea Lines of Communication, must clear in 24 hours
- USS Coronado tasked to clear the area of mine-like objects within 24 hours
 - USS Coronado's USV/UUV assets (fictional "JLSCS" automated surface interdiction) have maintenance problems that will delay their deployment
 - USS Coronado discovers UAV asset (fictional "RQ-X" autonomous airborne interdiction) available from USS San Diego, determines it can provide fill-gap capability, prepares plan, communicates with San Diego, transfers control, and returns
- Automated systems must consider user abilities, accuracy of data, changing political situation, etc.

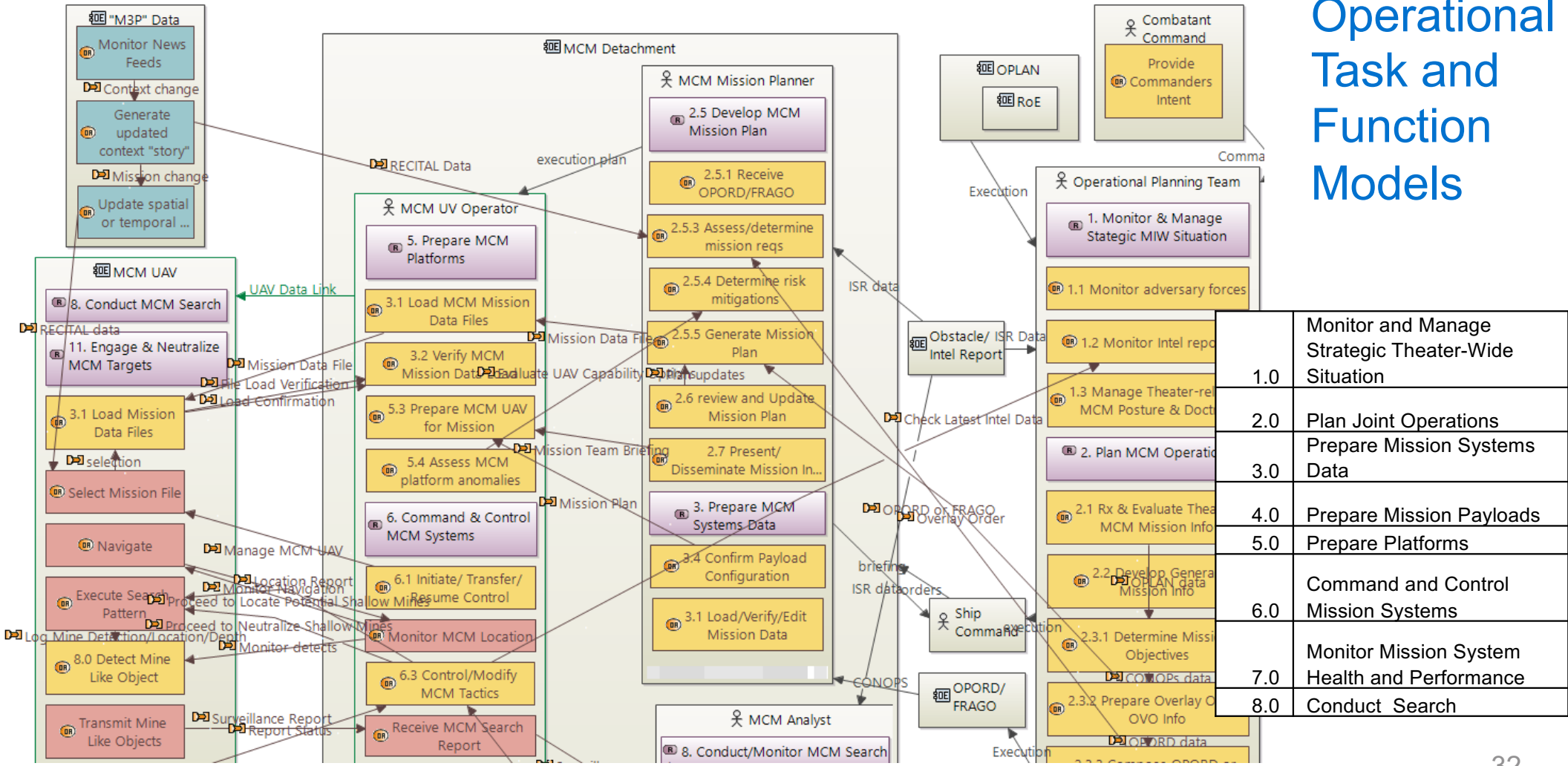
Task Analysis

1.0	Monitor and Manage Strategic Theater-Wide Situation
2.0	Plan Joint Operations
3.0	Prepare Mission Systems Data
4.0	Prepare Mission Payloads
5.0	Prepare Platforms
6.0	Command and Control Mission Systems
7.0	Monitor Mission System Health and Performance
8.0	Conduct Search
9.0	Assess and Classify Objects of Interest
10.0	Monitor Tactical Situation
11.0	Engage and Neutralize Targets of Interest
12.0	Record Mission Data
13.0	Perform Post-mission Activities
14.0	Analyze Mission Data
15.0	Conduct Mission Training and Exercises
16.0	Create and Maintain Digital Systems Models

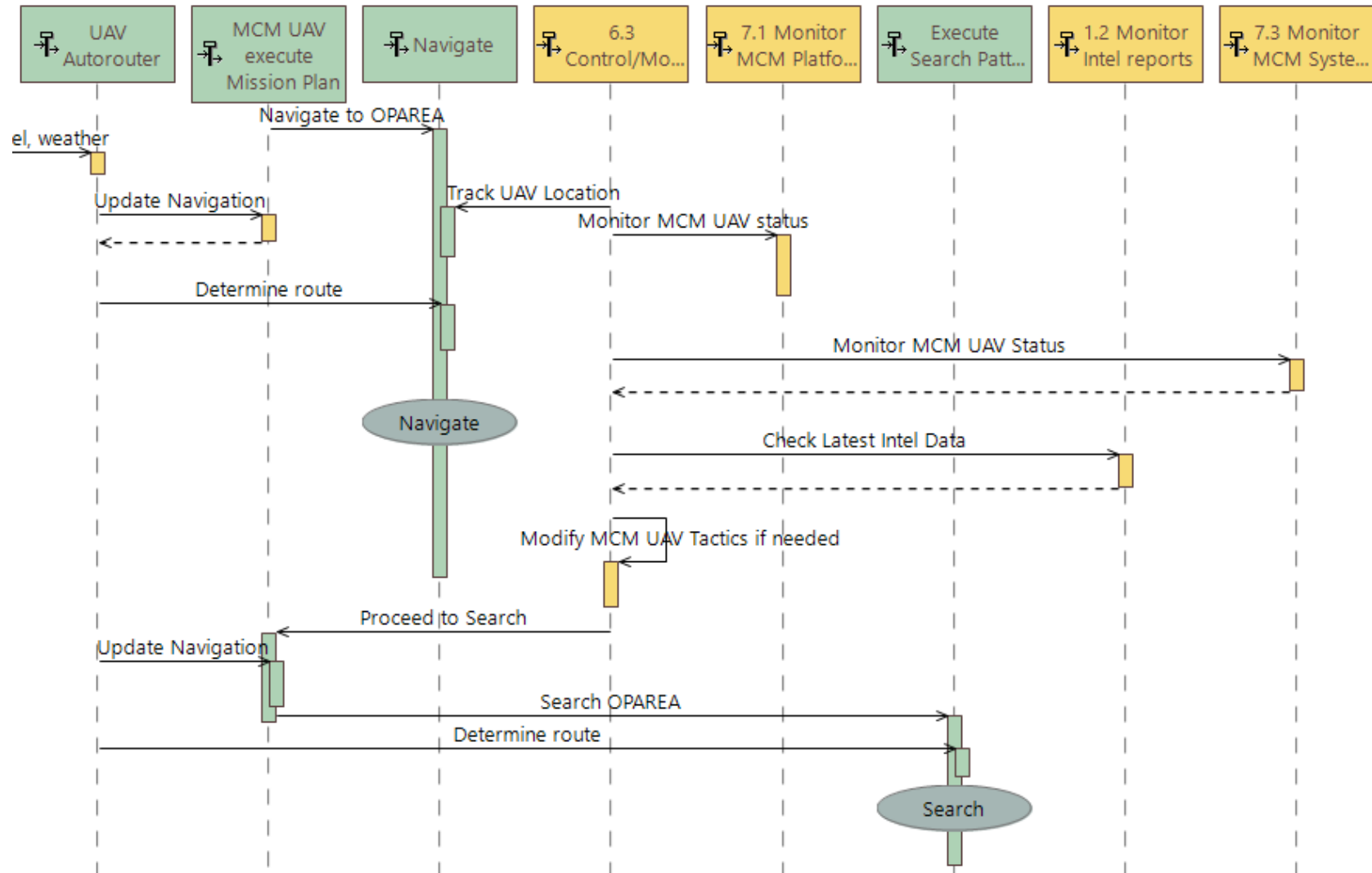
Primary Human Tasks

S-F	S-SF	Task	Function/Task Name
1.3			Manage theater security posture and doctrine
		b.	Develop and disseminate theater operations plans (OPLAN)
		d.	Prepare and disseminate information on theater Rules of Engagement
2.1			Receive and Evaluate Theater-wide Mission Information
		b.	Review strategic military objectives and theater Commander's intent
2.2			Develop General Mission Information
		a.	Review theater OPLANS
		c.	Determine and state mission objectives
		d.	Review ROE and other constraints
		f.	Identify and describe acceptable risks
2.3			Develop and disseminate operational orders (OPORD)
	2.3.3		Compose OPORD
		c.	Describe execution including commander's intent
2.5			Develop operational mission plan
	2.5.3		Assess mission requirements
		c.	Assess sensor and weapon requirements
		d.	Assess C3 requirements
		e.	Identify available assets
		f.	Select assets for mission
	2.5.4		Determine risk mitigation methods
		a.	Assess threat-related risks
		b.	Assess environment-related risks
		c.	Assess malfunction-related risks
		d.	Develop plans to mitigate risks

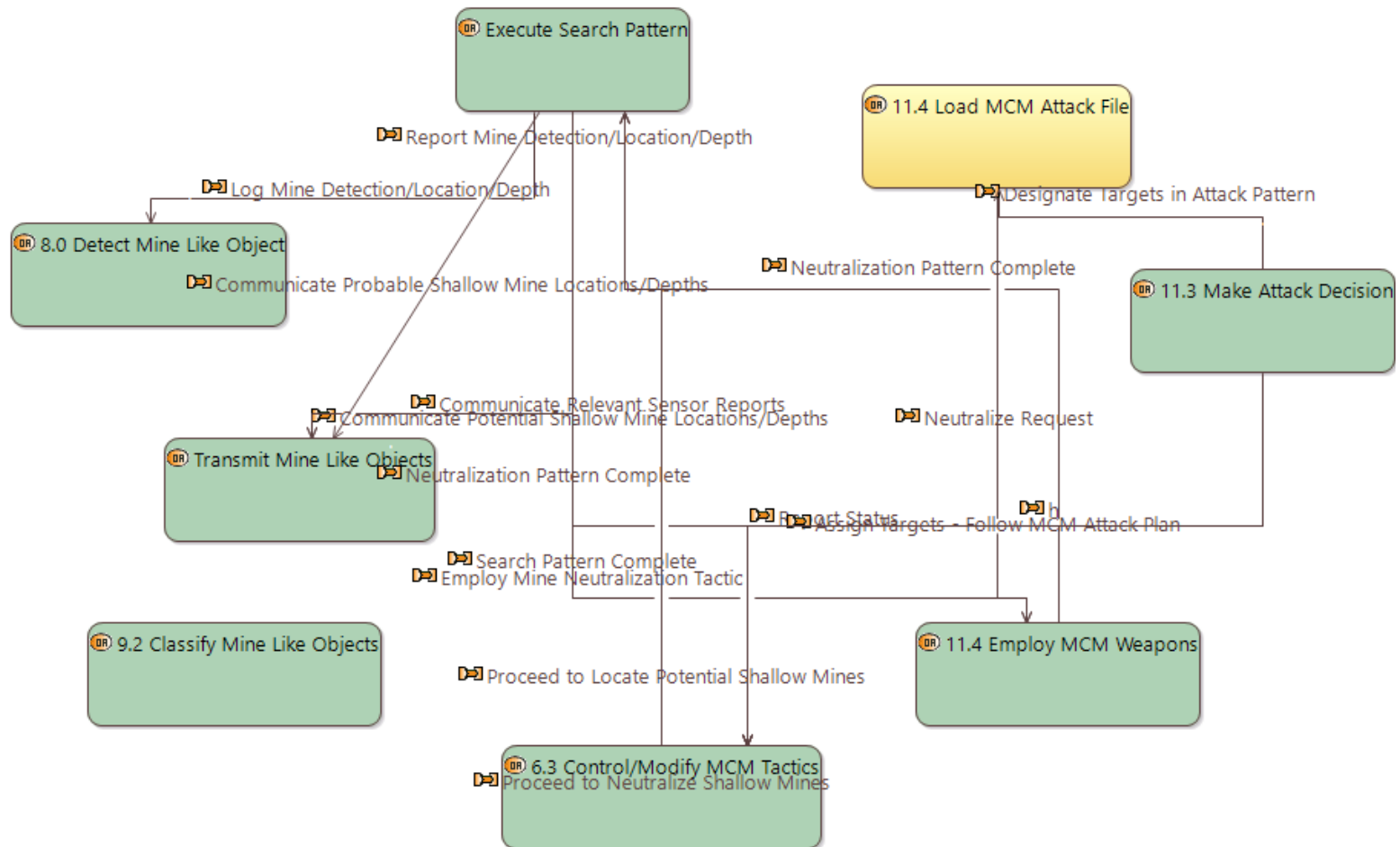
Combined Operational Task and Function Models



Operational Activity Model

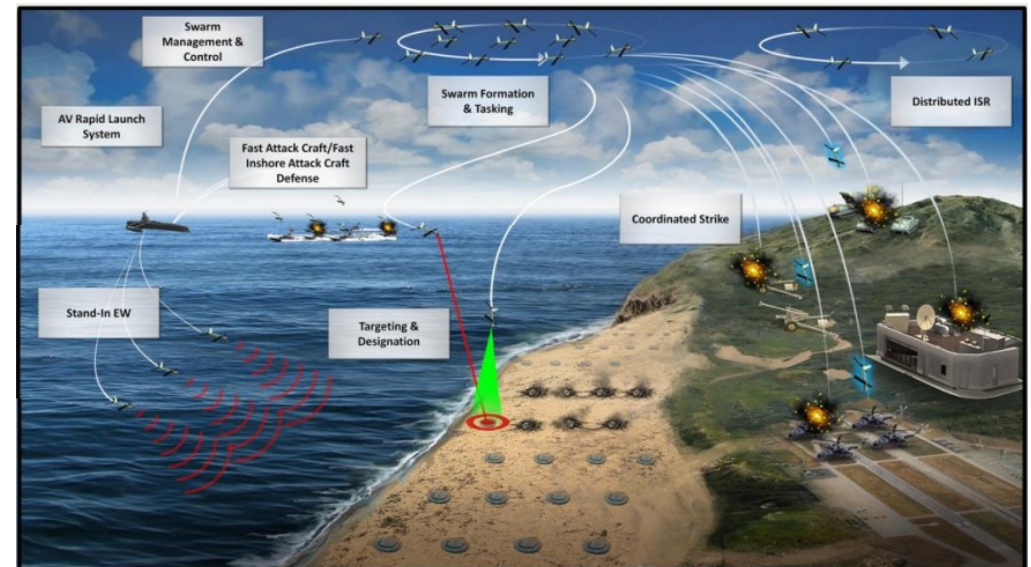


Automated System State Model for Countermine Activities



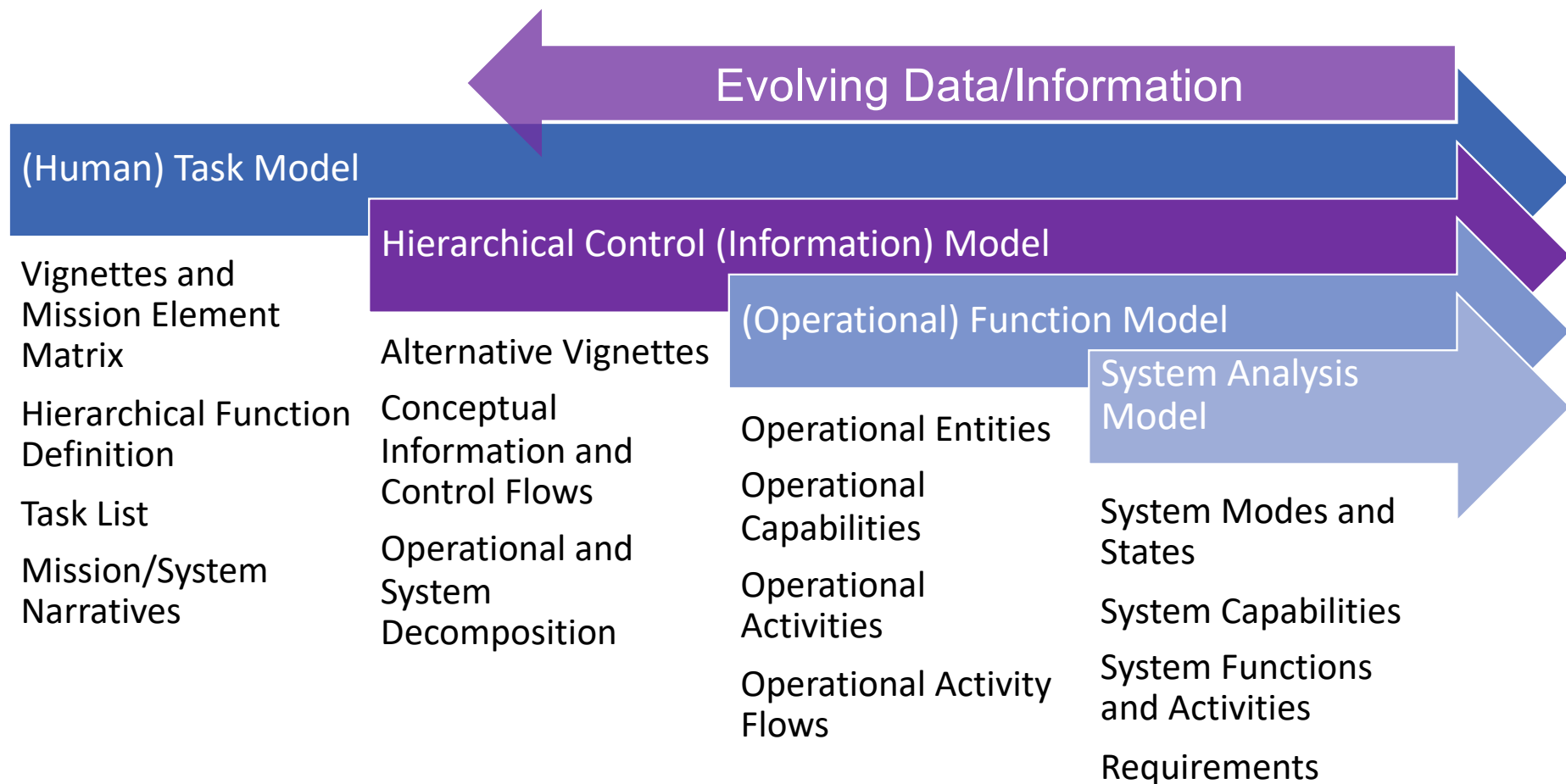
Hierarchical Control of Distributed Autonomous Human-Machine Teams

- Stochastic decision processes
- Controlled by both machine agents and humans
- Ideally leverage the distinct capabilities of each
- Must address the challenge of transferring control quickly, safely, and smoothly back-and-forth between the agent and the human
- Can be viewed as hierarchical levels of control using non-hierarchical distribution of information



Office of Naval Research, Code 30 overview briefing

Expanded modeling flow





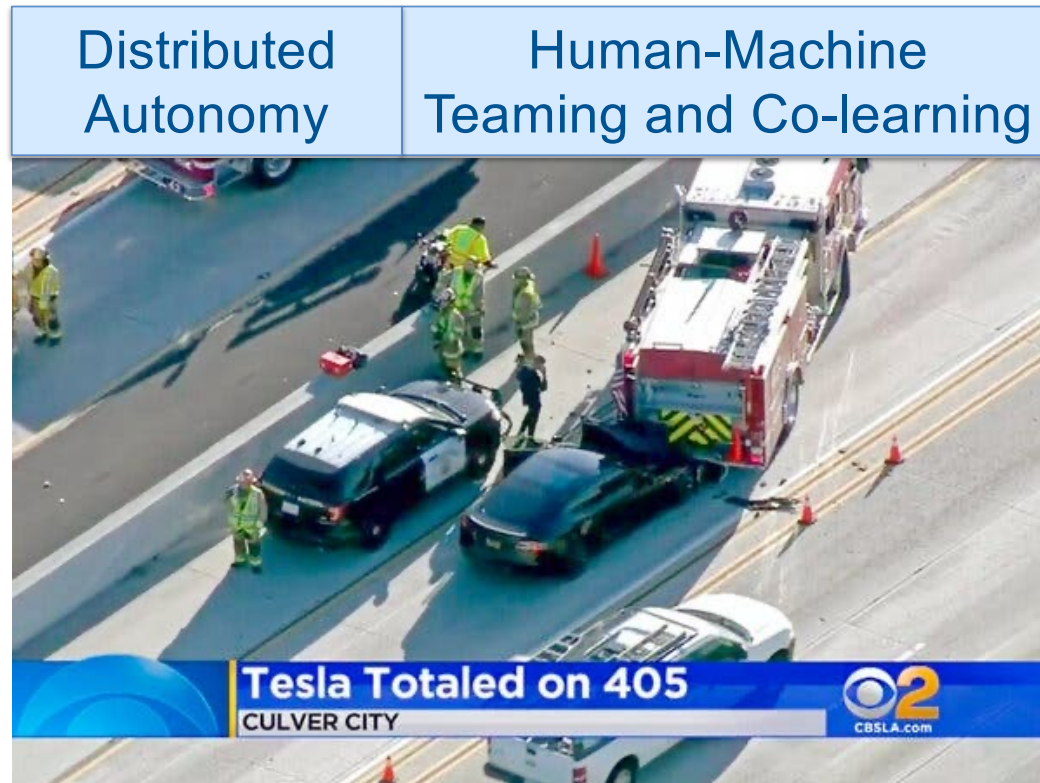
Complaint claims Tesla's 'Full Self-Driving' software caused crash

14 November 2021

US safety regulator opens investigation into Tesla Autopilot following crashes with parked emergency vehicles



U.S. auto regulators have opened a preliminary investigation into Tesla's Autopilot advanced driver assistance system, citing 11 incidents in which vehicles crashed into parked first responder vehicles while the system was engaged. The Tesla vehicles involved in the collisions were confirmed to have either have had engaged Autopilot or a feature called Traffic Aware Cruise ... Continue reading




NY Times photo

There are 9.1 driverless car crashes per million miles driven. Regular vehicles have a rate of 4.1 crashes per million miles driven. Fewer severe injuries are caused by self-driving cars.

(carsurance.net/insights/self-driving-car-statistics)

Transfer of Authority between human and machine remains a concern.

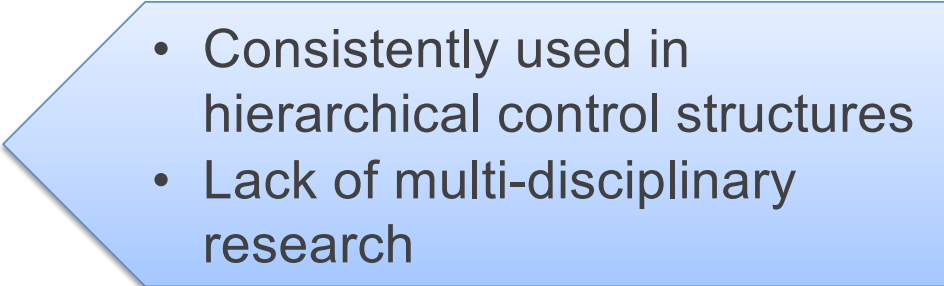
CONTROL AUTHORITIES

	ODI RESUME 	
	Investigation: EA22002 Prompted By: PE21020 Date Opened: 06/08/2022 Investigator: Steven Posada Approver: Tanya Topka Subject: Autopilot System Driver Controls	Date Closed: 04/25/2024 Reviewer: Gregory Magno
MANUFACTURER & PRODUCT INFORMATION		
Manufacturer:	Tesla, Inc.	
Products:	2012 – 2023 Model Y, X, S, 3 equipped w/ Autopilot manufactured up to 7-Dec-2023	
Population:	2,031,220	
Problem Description:	The prominence and scope of Autopilot's control may be insufficient to prevent crashes due to lack of driver engagement.	

“Of the remaining 467 crashes, ODI identified trends resulting in three categories: collisions in which the frontal plane of the Tesla struck another vehicle or obstacle with adequate time for an attentive driver to respond to avoid or mitigate the crash (211), roadway departures where Autosteer was inadvertently disengaged by the driver’s inputs (111), and roadway departures in low traction conditions such as wet roadways (145). ODI observed this pattern across all Tesla models and hardware versions. Crash and human factors assessment showed that Autopilot controls did not sufficiently ensure driver attention and appropriate use. At the same time, peer analysis and vehicle evaluations established that Autopilot invited greater driver confidence via its higher control authority and ease of engagement. This mismatch of weak usage controls and high control authority was evident in these crash categories, which included indications of driver disengagement from the driving task. This mismatch was also evident in roadway departures when the system was engaged in low traction conditions outside of Tesla’s recommendations.”

A Model for the Information Model

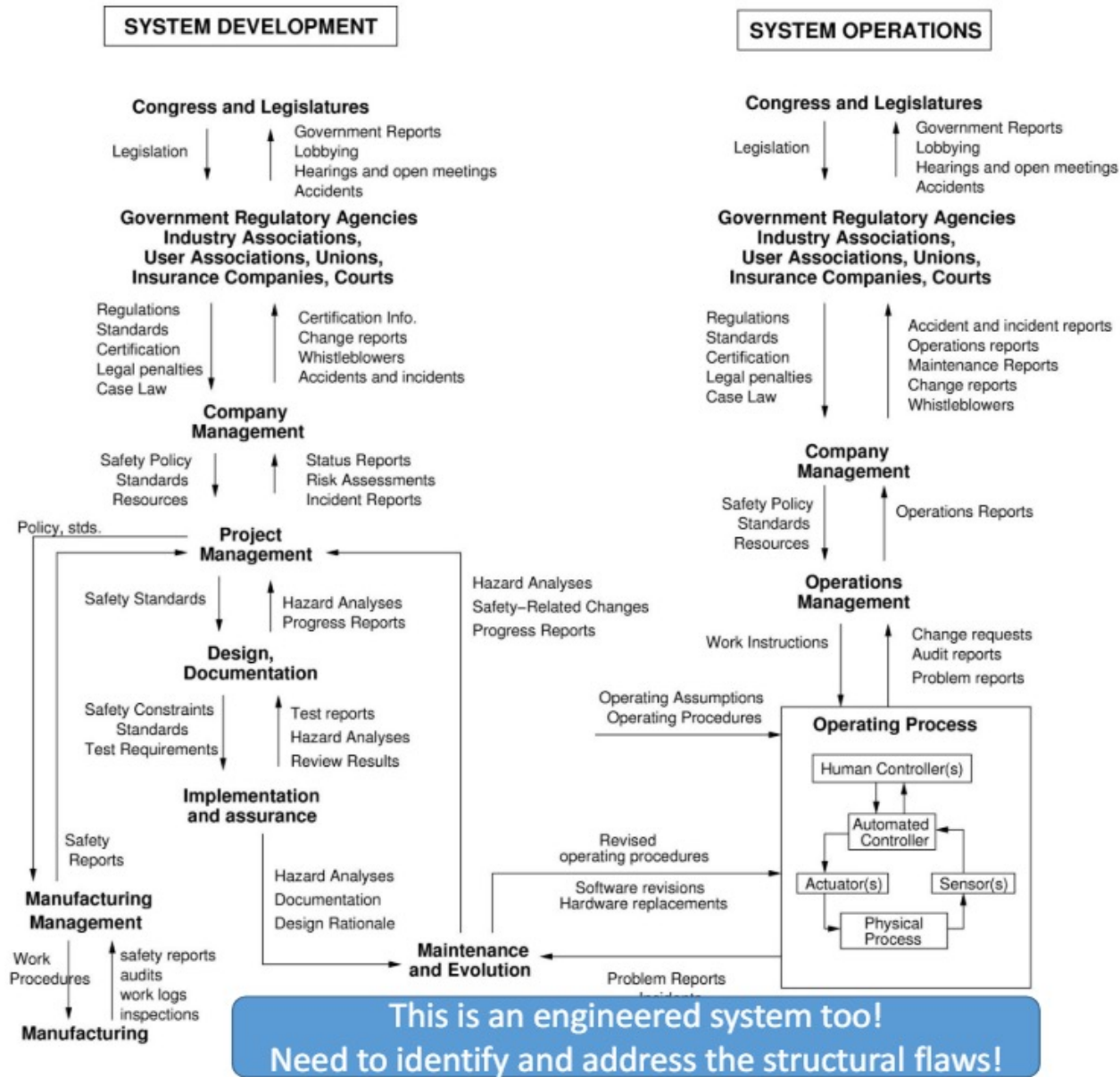
- Most accidents/mission failures will be caused by errors in **interpretation of information** by either the human or the machine
- Leading to errors **transfer of control (or authority)** made in the planning process and instantiated in the live situation
- Underlying concept of human informational transfer has subjectivity
 - **Intent**
 - **Rules**
 - **Authorities**
 - **Other Contextual Information**
- Desire a Systems Engineering approach to address both information design and control mechanization across layers of hierarchy
- Rigorous approach defined in Leveson's STAMP/STPA methodology

- 
- Consistently used in hierarchical control structures
 - Lack of multi-disciplinary research

Example Safety Control Structure

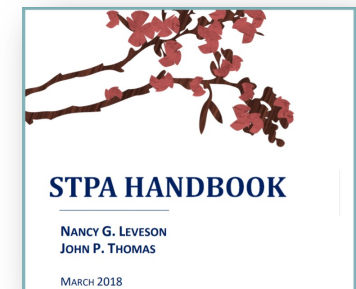


(Leveson, 2012)

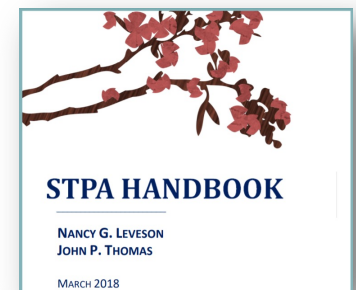
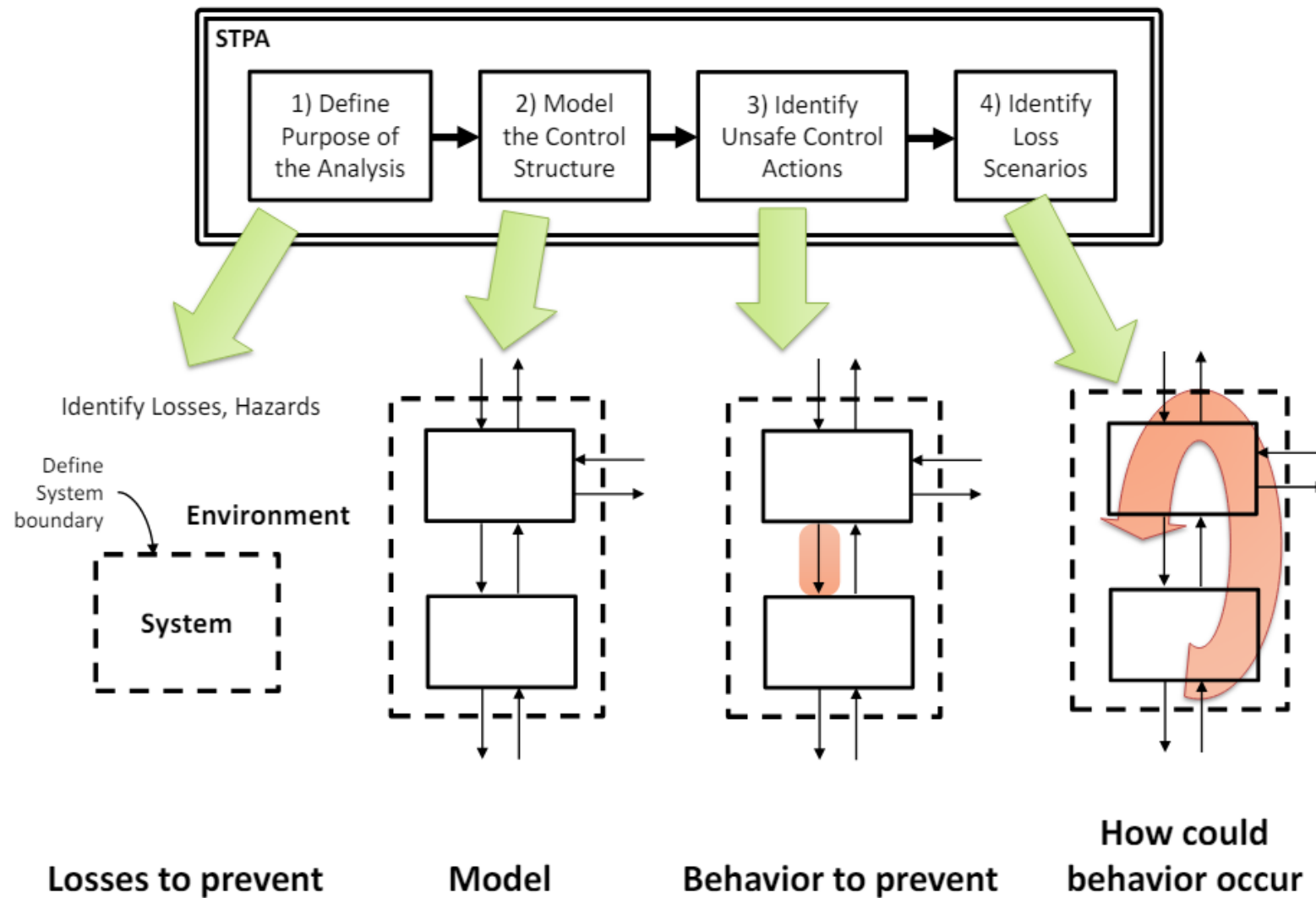


STAMP-STPA

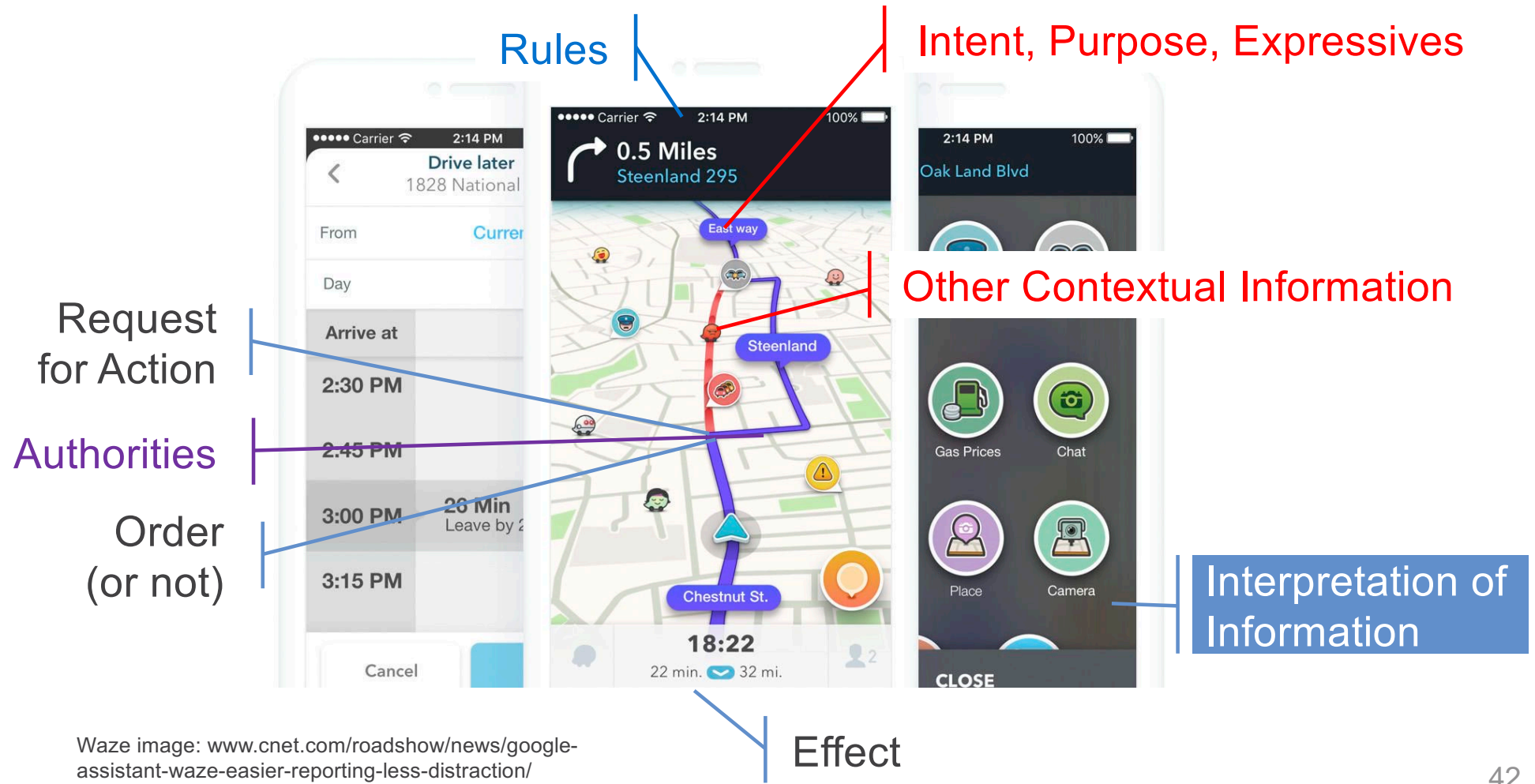
- Intent
- Rules
- Authorities
- Other Contextual Information



STPA process

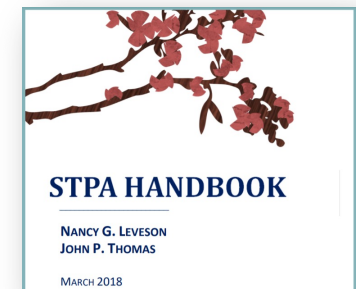
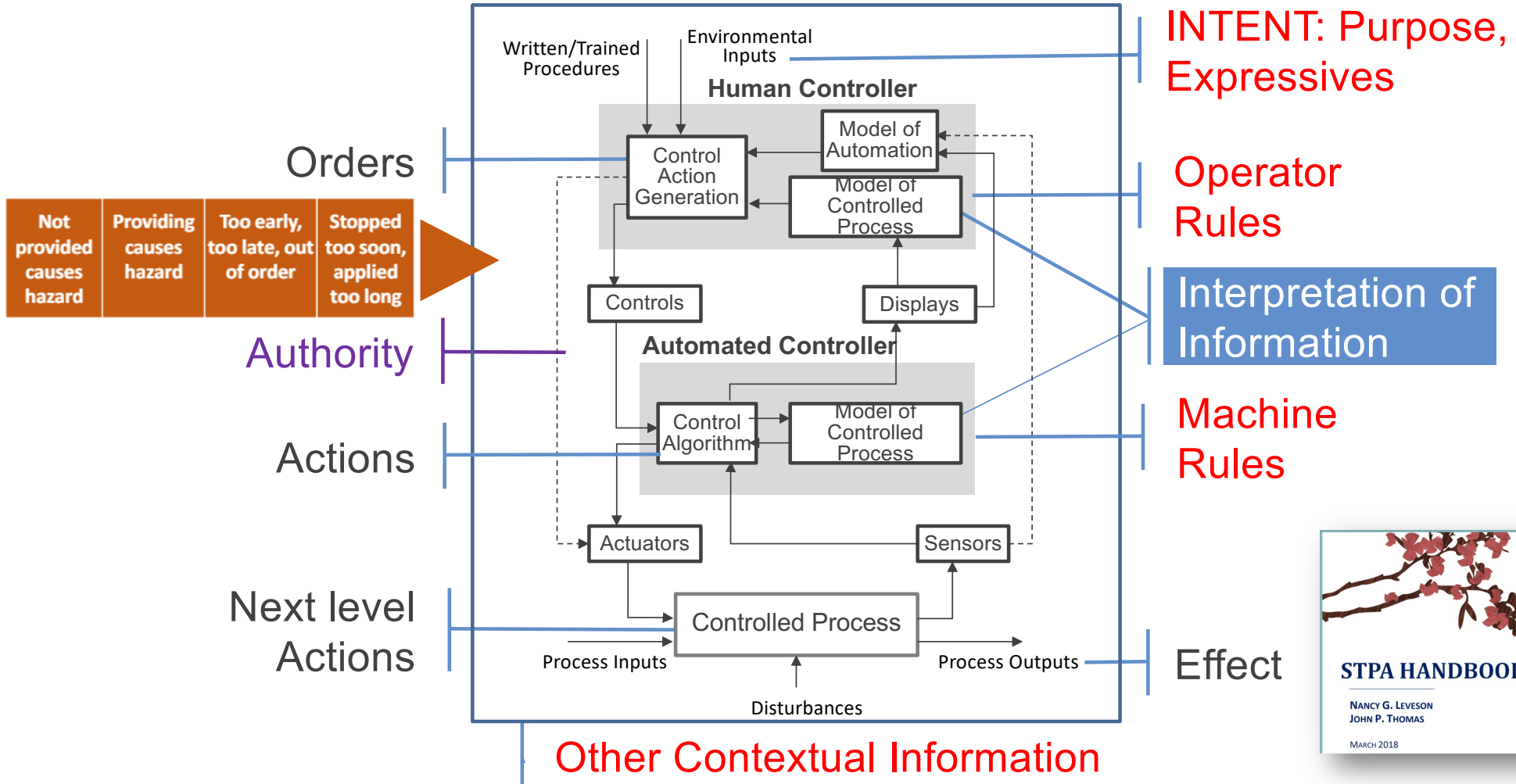


WAZE Human-Machine Teaming

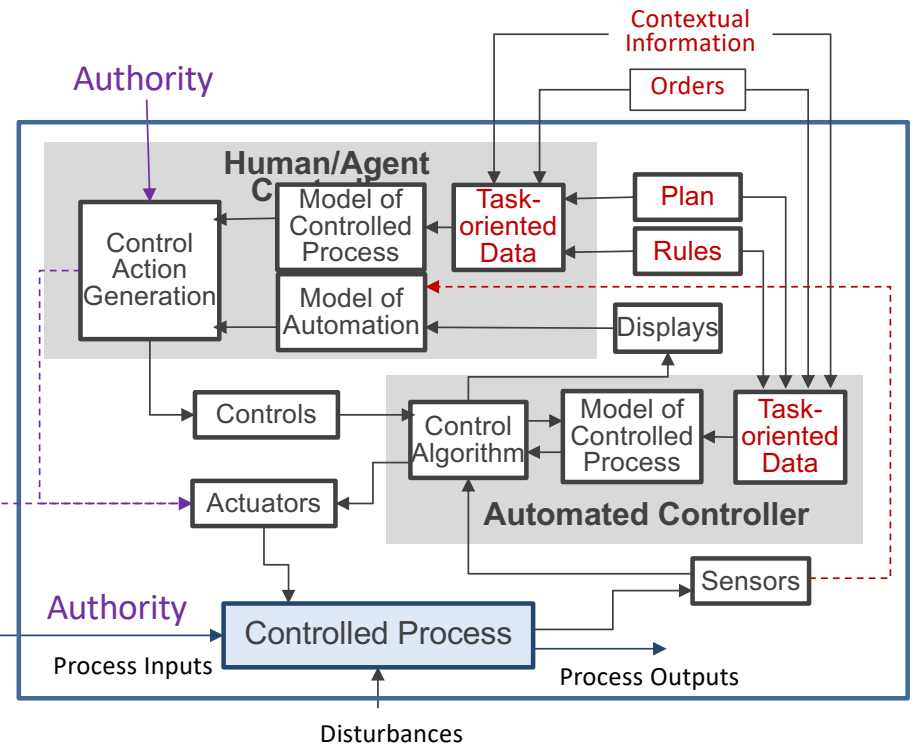
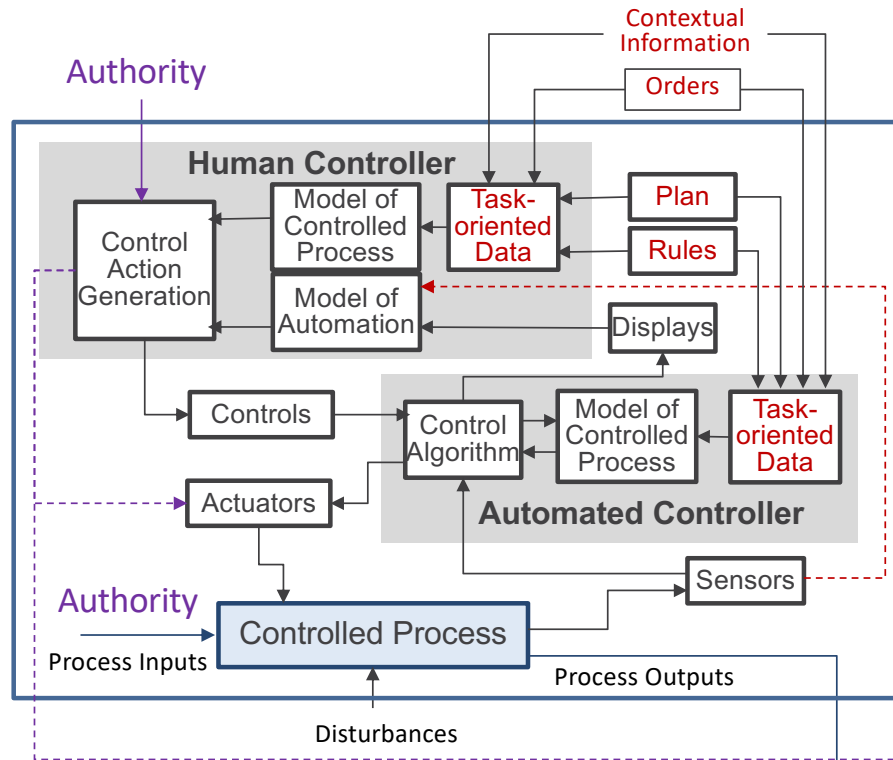


Waze image: www.cnet.com/roadshow/news/google-assistant-waze-easier-reporting-less-distraction/

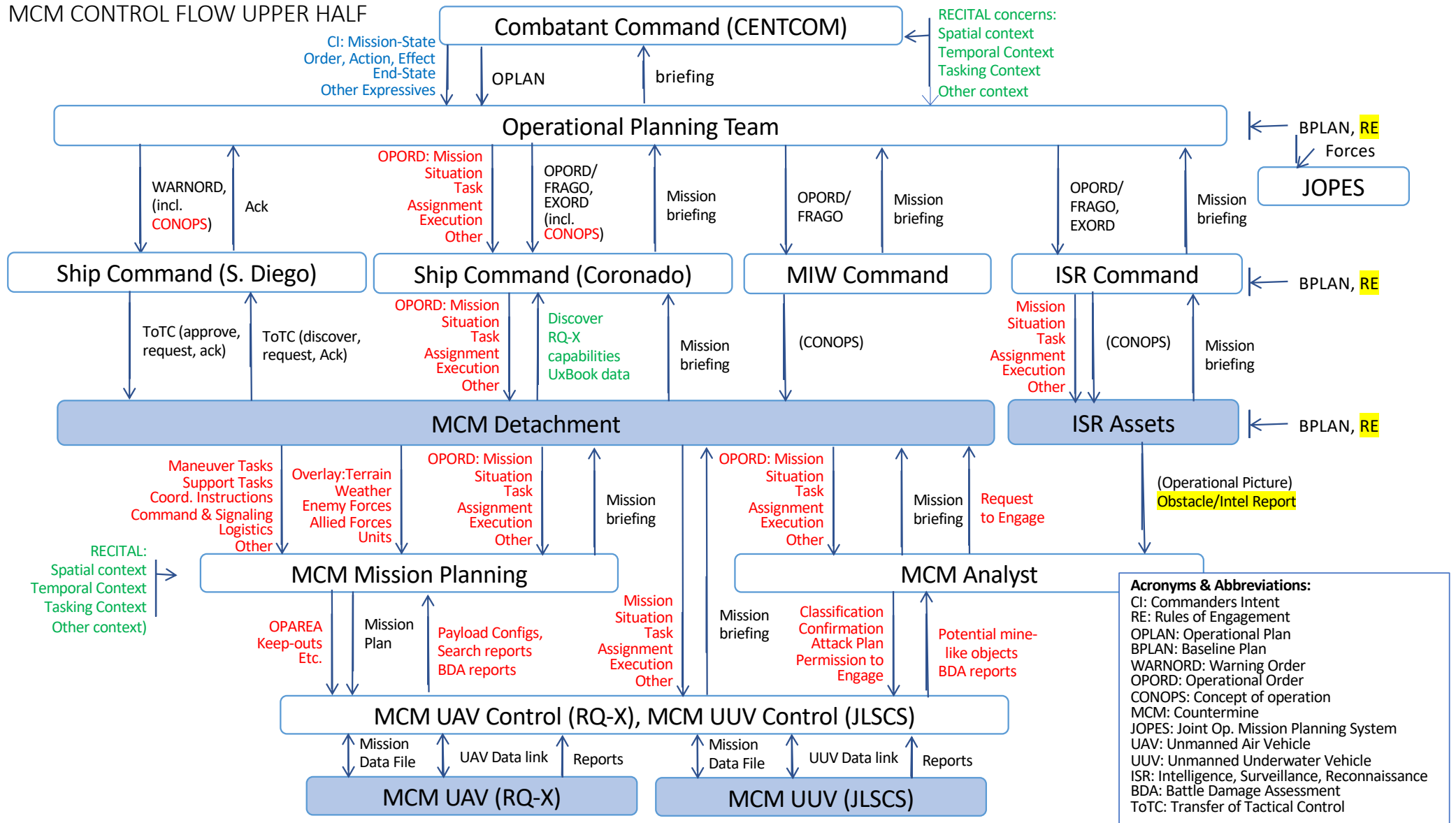
Controller Model



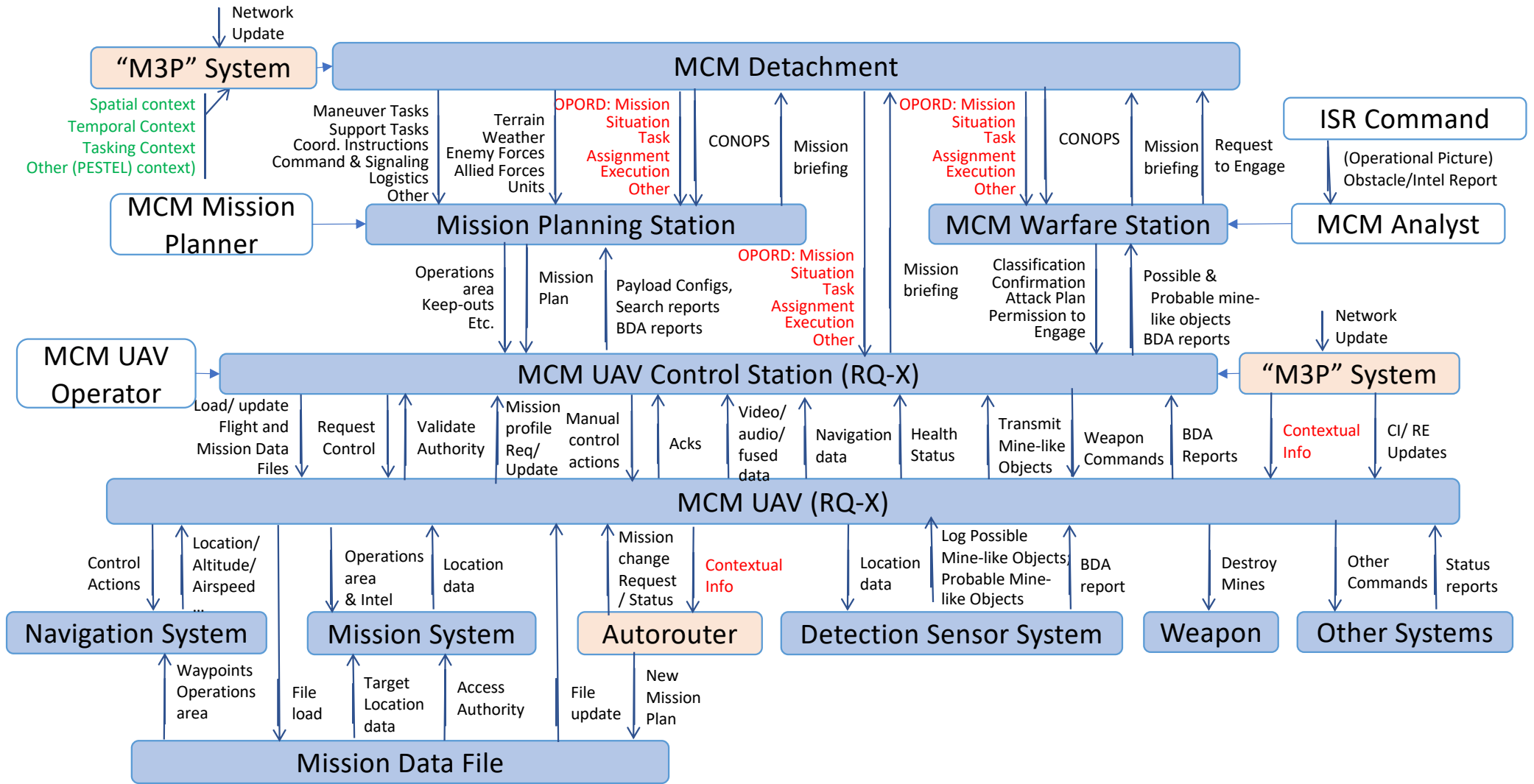
STPA-based Information Model



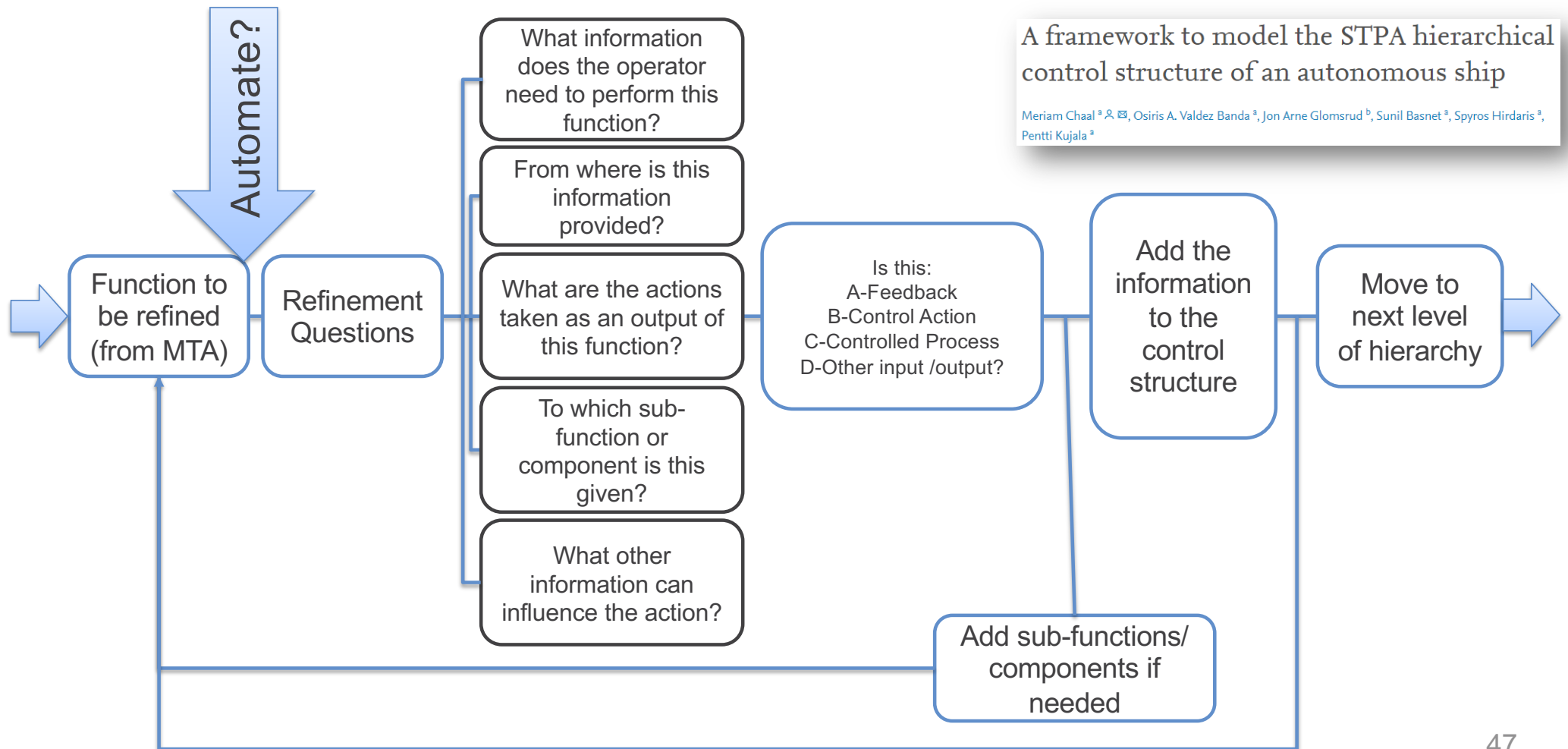
MCM CONTROL FLOW UPPER HALF



UAV MCM CONTROL FLOW LOWER HALF



Task to Function Decomposition Process



Example Functional Analysis for an automated operator

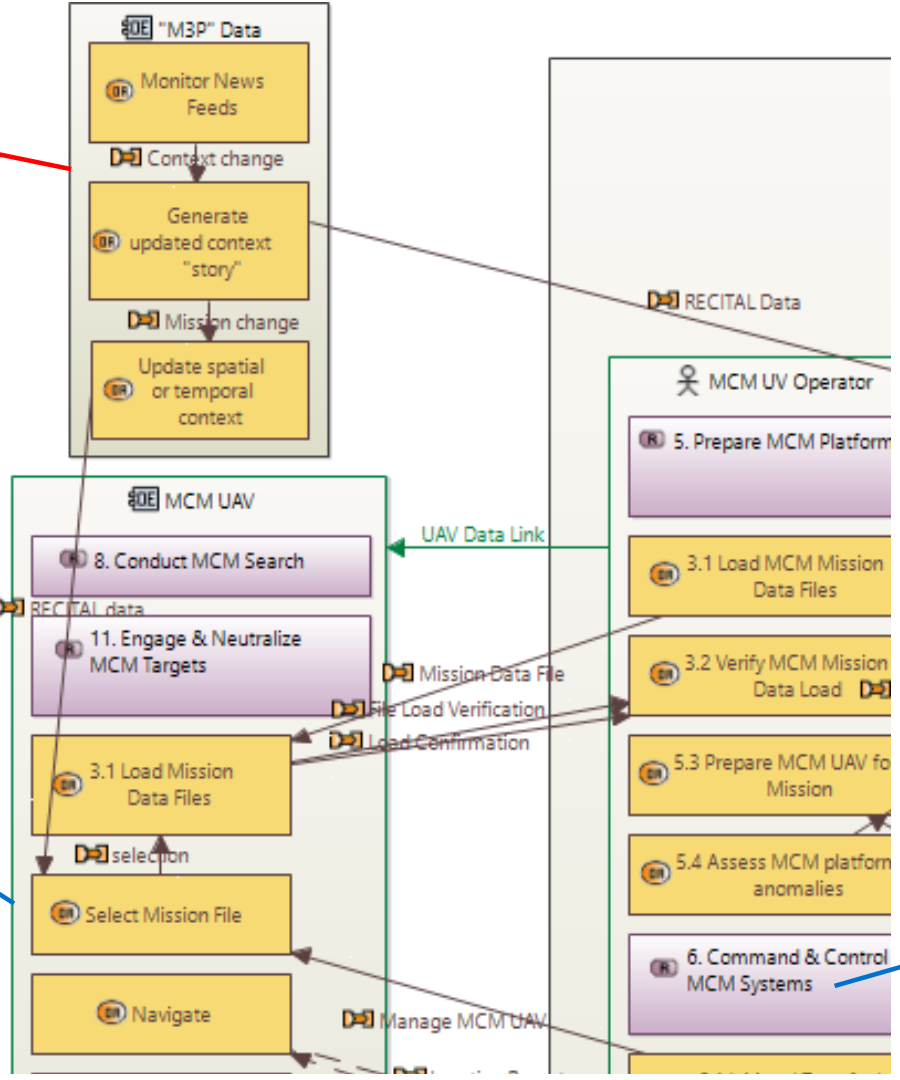
Function	What does the operator need to perform this function?	From where is this information provided?	What are the actions taken as an output to this function?	to which other function is it given?	What other information can influence the action?	comments
8.X Navigate	Navigation data; location altitude, airspeed, etc.	GPS and INS; UAV navigation system	Navigate to next waypoint, or modify MCM tactics	6. MCM UAV Operator Control/ Modify MCM Tactics	Higher commands to 6.2 Interrupt MCM Platform Execution or 6.4 Scuttle MCM Systems	<i>the UAV could take its own action to alter navigation based on RECITAL data?</i>
	MCM mission file w/ OPAREA & possible mine-like objects	loaded MCM Mission File	Update MCM Mission File	6. MCM UAV Operator Control/ Modify MCM Tactics	Change to Mission Plan, 3.3 Edit MCM Mission Data Files	
8.X Contextual Auto-routing	Contextual information that would instigate a change in navigation: weather, intel changes, shift in OPAREA; as well as MCM UAV knowledge of its current statuses	various operations centers are monitoring changing conditions and dynamically adjusting potential routing choices (new M3P scenario)	The MCM makes a decision to change its navigation based on selection of new routing alternatives as determined by the input data	Navigate; 6. MCM UAV Operator Control/ Modify MCM Tactics	Alternate instructions from the MCM UAV Operator; Concern about an unauthorized source?	<i>In this function, the MCM UAV can self-initiate a change to its routing based on an external context change (like Waze recommending a switch to an alternate route)</i>

Operational Task Model

Contextual information that would instigate a change in navigation: weather, intel changes, shift in OPAREA; as well as MCM UAV knowledge of its current statuses

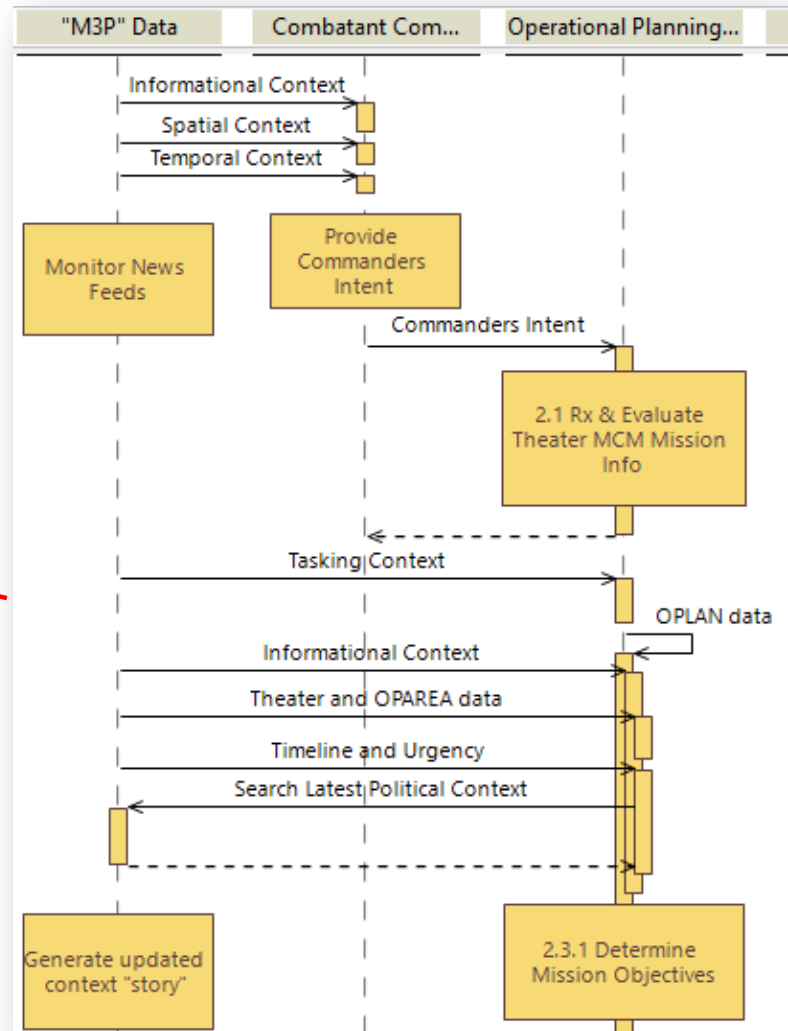
The MCM makes a decision to change its navigation based on selection of new routing alternatives as determined by the input data

In this function, the MCM UAV can self-initiate a change to its routing based on an external context change



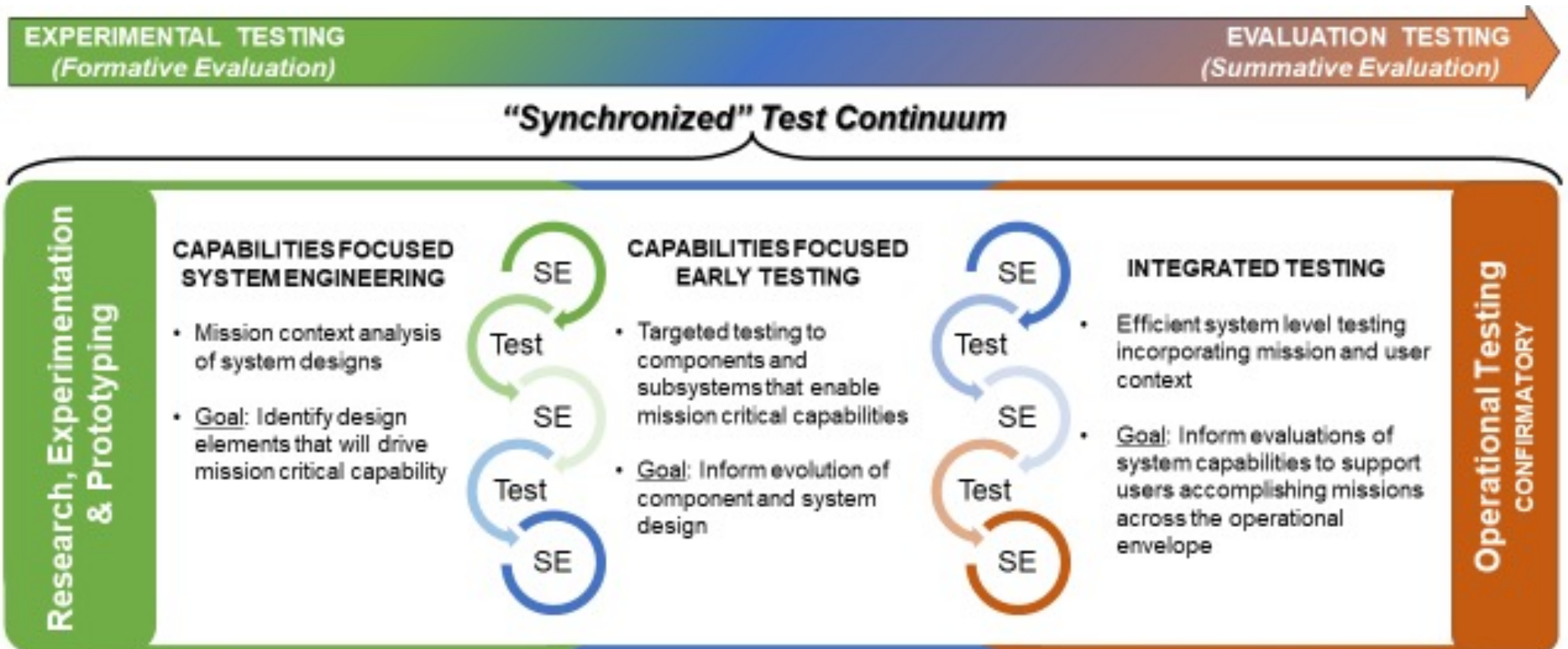
Navigate; 6. MCM UAV Operator Control/ Modify MCM Tactics

Operational Information-driven Activity Model

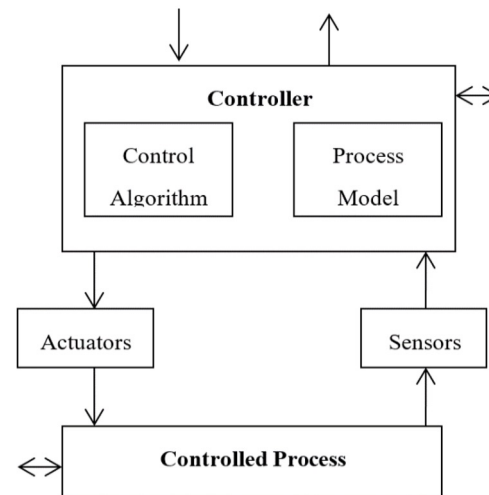
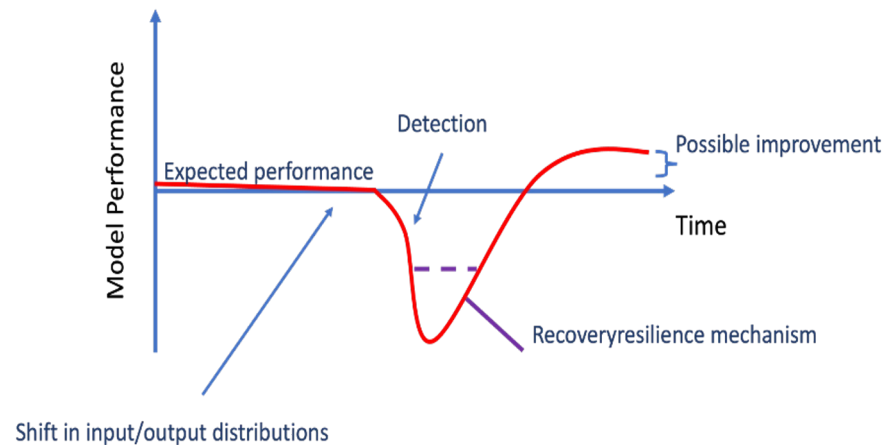
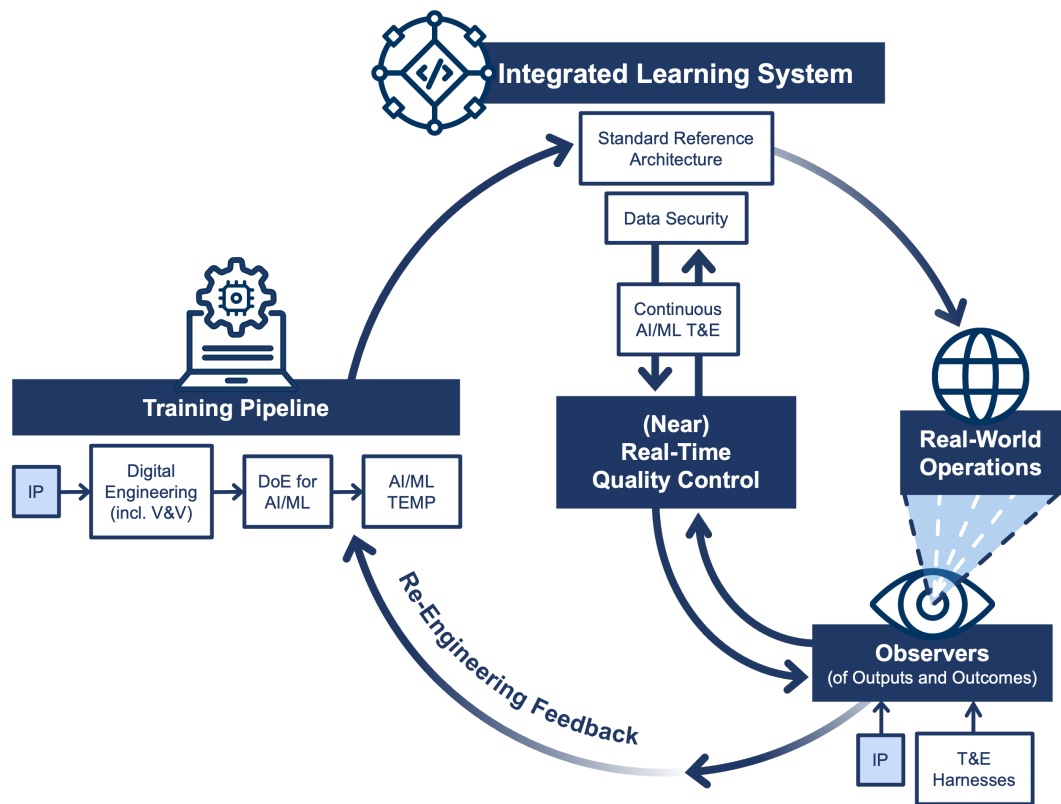


Contextual information that would instigate a change in autonomous UAV operation: weather, intel changes, shift in OPAREA; as well as UAV knowledge of its current statuses

1. SE4AI and AI4SE and the SERC Research Roadmap
2. Systems Engineering and AI
3. **Human-Machine Teaming**
 1. Building user Trust by understanding the Human AI system
 2. Architecting AI Systems for long-term trust: Linking task & function allocation, test and risk analysis and need for systems testbeds
 3. T&E as a Continuum: what to test and how to interpret for AI Systems of varying complexity and embeddedness
 4. AI Resilience : Strategies to mitigate disruptions / ensure acceptable behaviors and recoveries when failures occur



FRAMEWORK FOR AI RESILIENCE THROUGH EVALUATION OF SYSTEMS AND TECHNOLOGY (FAIREST)



QUESTIONS?