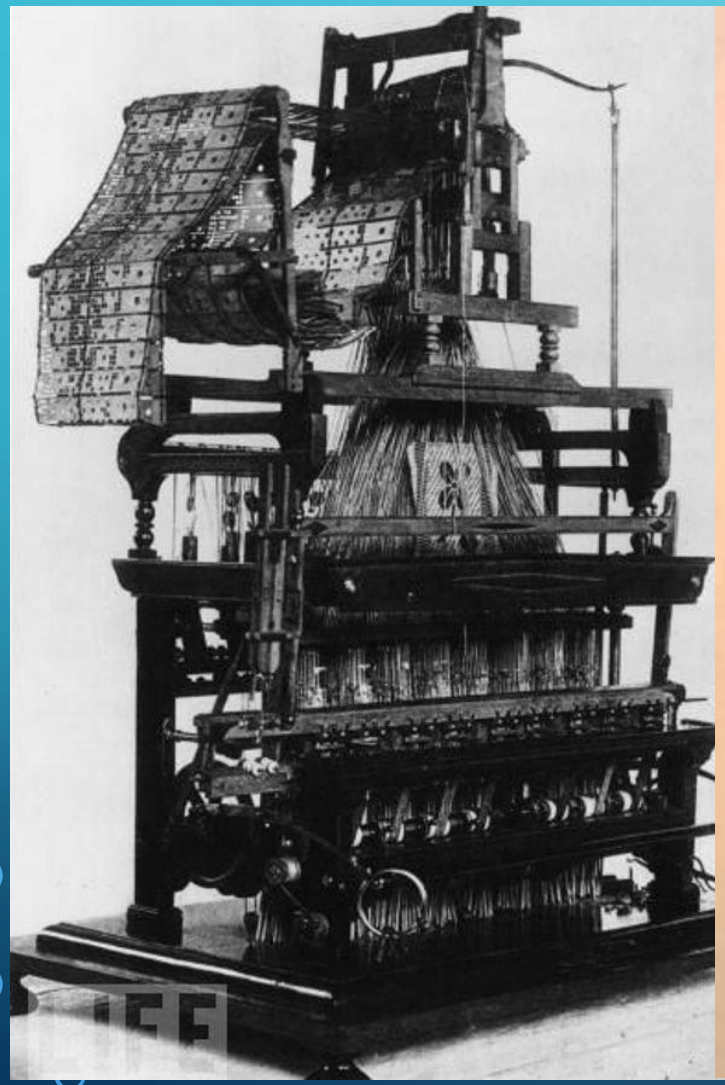
ASSURING HUMAN CONTROL AUTHORITY OVER AUTONOMOUS SYSTEMS

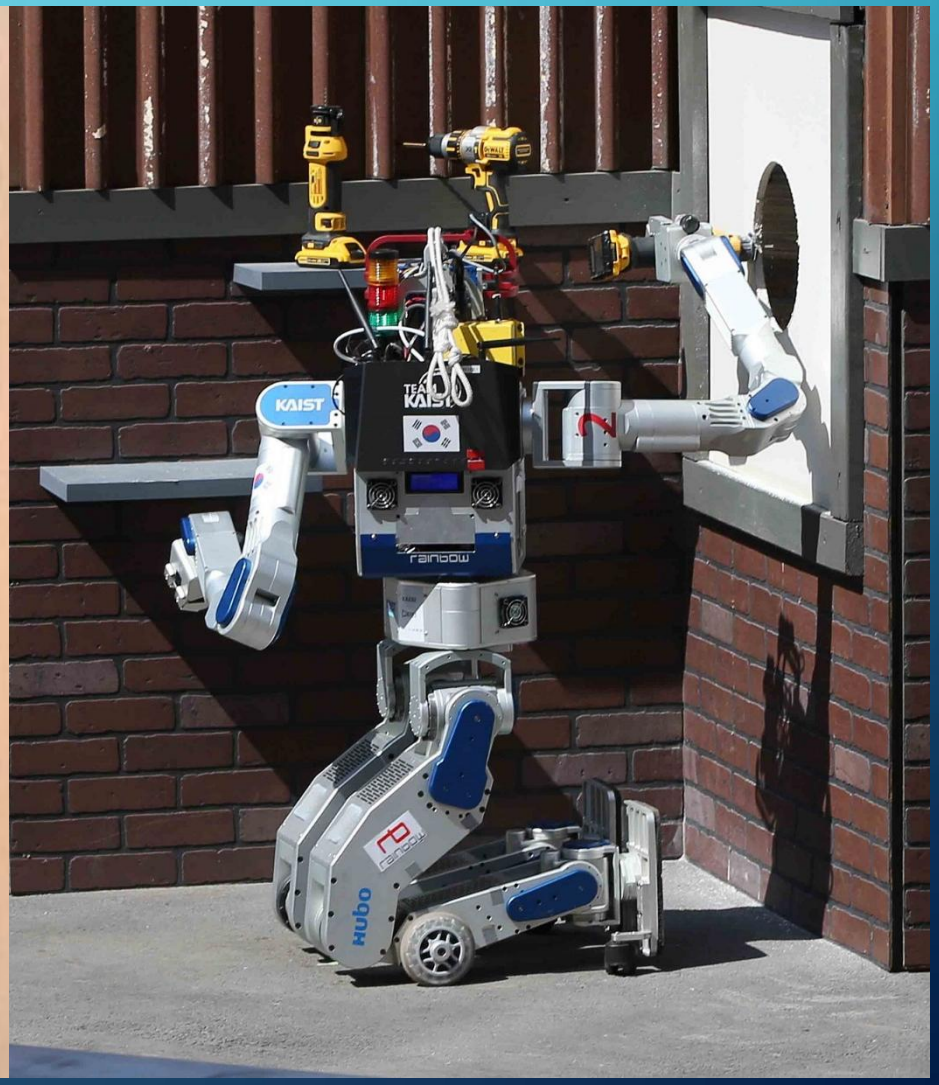
JEN NARKEVICIUS, PHD JENIUS LLC

AND

STEVE HARRIS, RATIONAL LLC

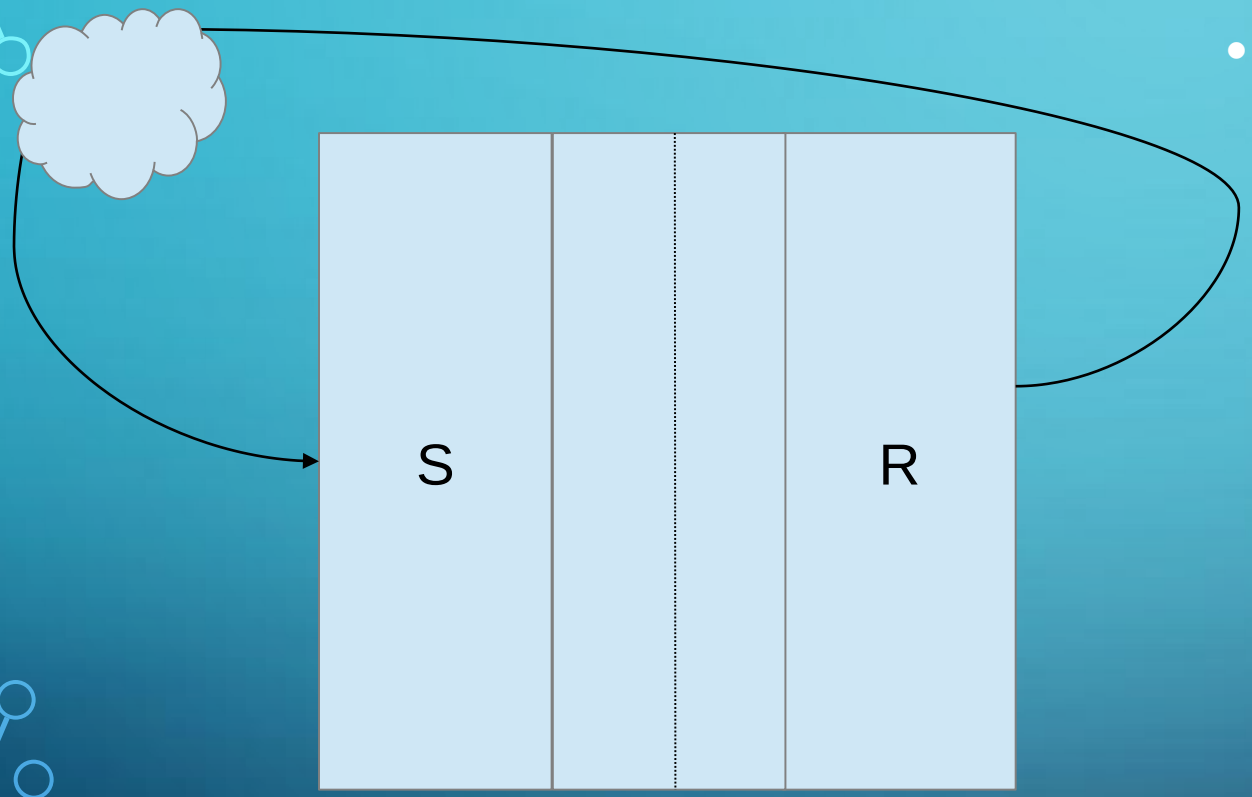
AUTOMATION OVER TIME





http://www.npr.org/sections/pictureshow/2011/04/01/135018381/on-apples-birthday-a-brief-history-of-computing
http://erricous.free.fr/Electronique/Projets/thermost/THER005.JPG
https://irf.kaist.ac.kr/board/newsView.do?guid=c922692d-870d-e511-9407-2c44fd7df8b9

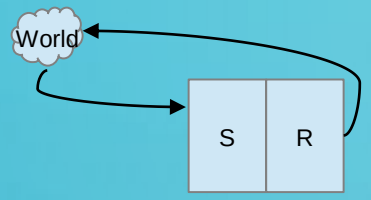
HUMAN-SYSTEM FUNCTION ALLOCATION

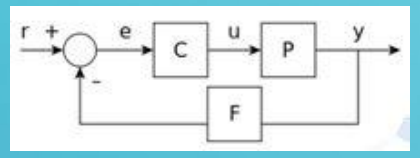


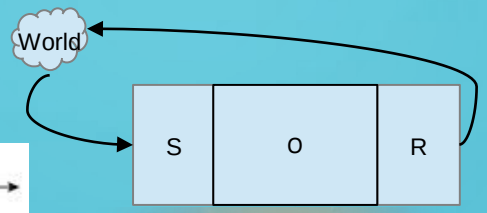
- Note that we have not specified the human or machine components in this depiction of closed-loop control, only the functional requirements
  - induction and deduction
  - internal representation (simulation)
  - estimates of missing parameters

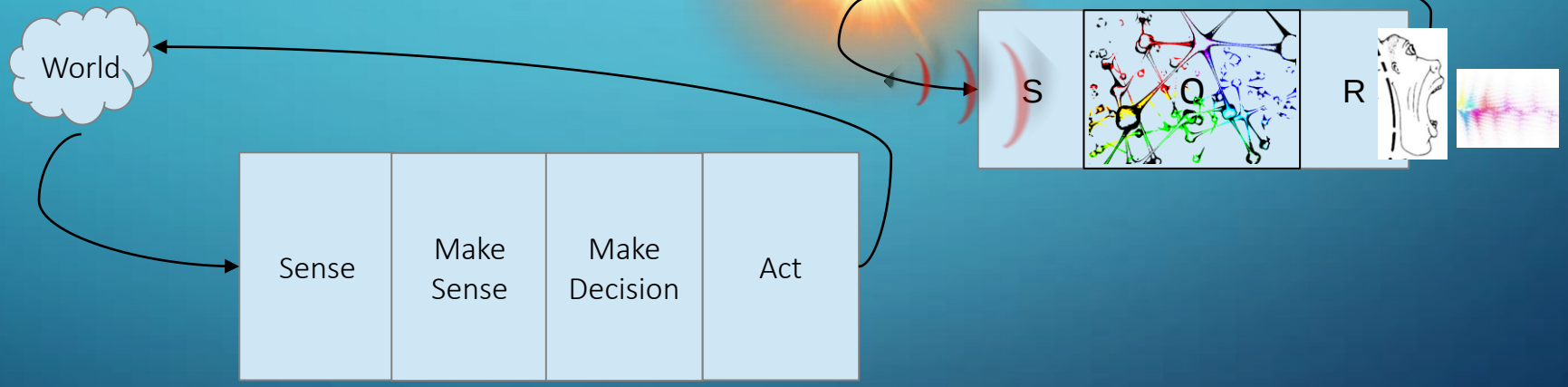
Ref: Emergent Failure Modes and What to Do about Them, Steve Harris & Jen Narkevicius, 26th Annual INCOSE International Symposium (IS 2016) Edinburgh, Scotland, UK, July 18-21, 2016

COMPLEXITY AND CONTROL



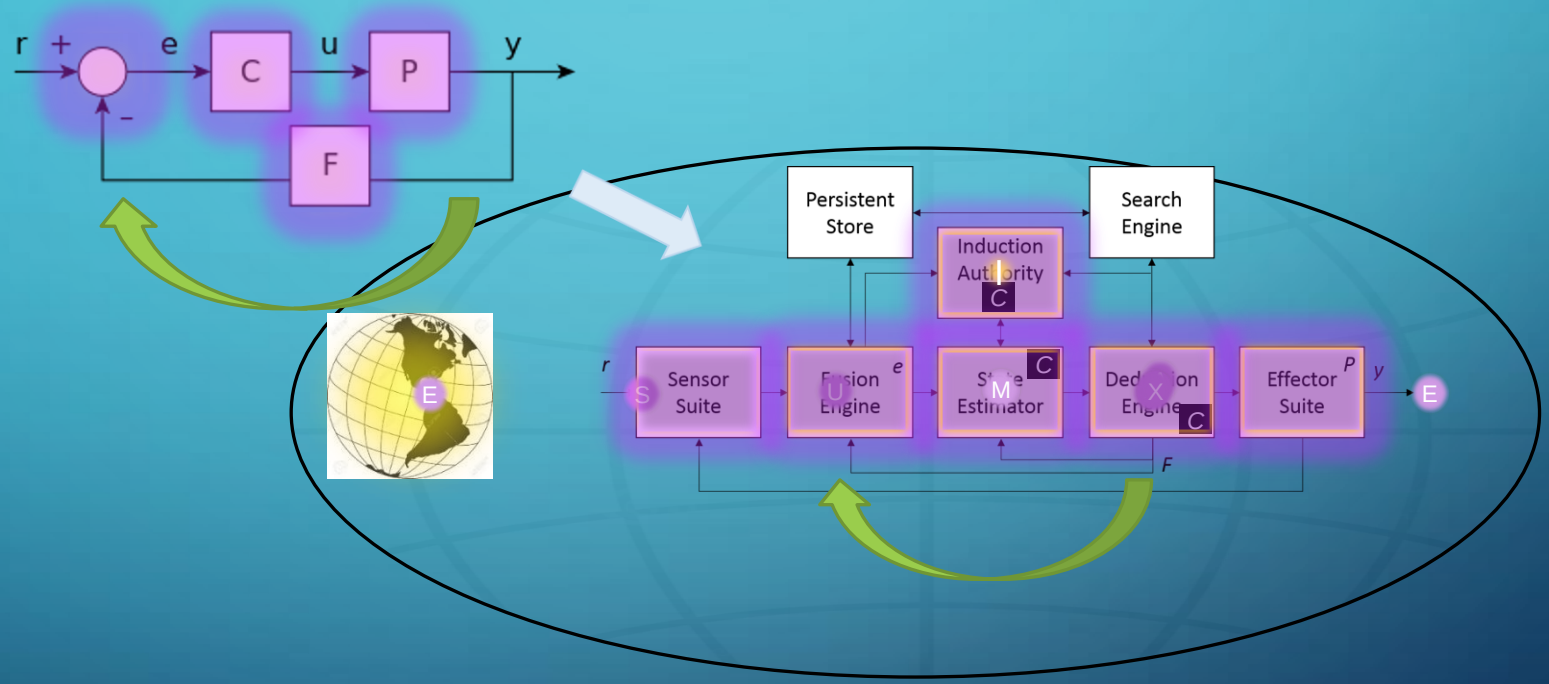






Process *taxonomy*

CONTROL THEORY



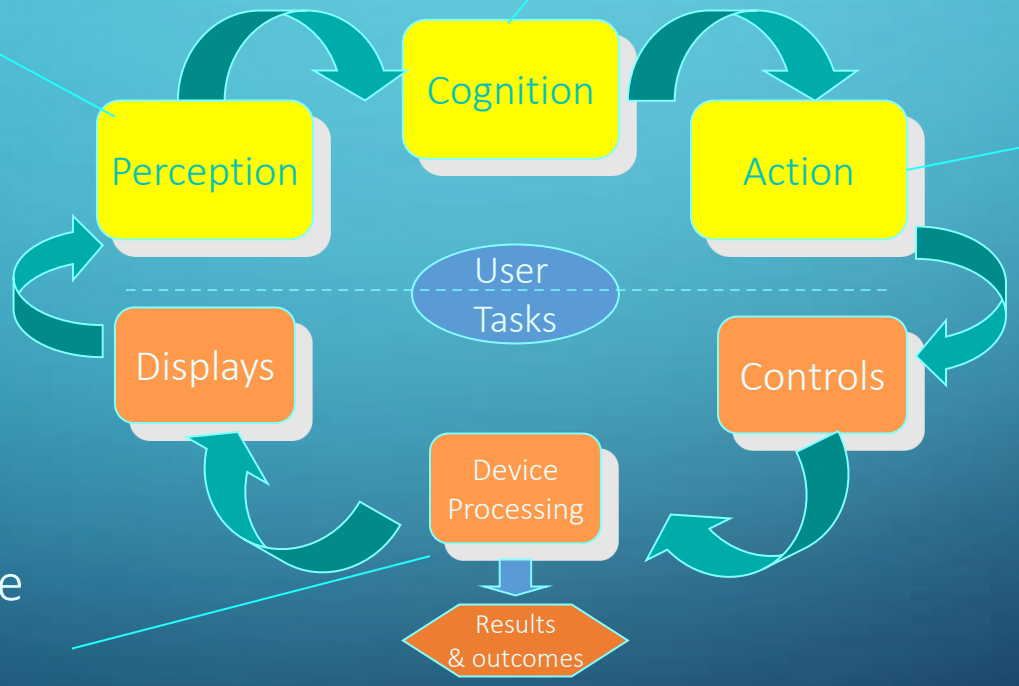
MODEL OF USER-DEVICE INTERACTION*

Events, incidents and alerts

Ways users organize their thinking about the system under control.

Information Management & Display

Help user gain situation awareness



Decision support

Help user understand what the options are for action.

Architecture Functional Requirements

*After diagram by Robert A North, PhD, Human-Centered Strategies, LLC, based on FDA UPCARE model

EMERGENT FAILURE MODES

http://www.noaanews.noaa.gov/stories/s2015.htm ; http://whotv.com/2014/09/26/all-chicago-flights-halted-by-fire/
https://www.google.com/search?q=super+bowl+power+outage&source=lnms&tbm=isch&sa=X&ved=0CAkQ_AUoA2oVChMI5MK7kLDRyAIVyCgeCh0TPwSX&biw=1366&bih=628&dpr=1.71#imgrc=XiOWIVhrFNqFsM%3A

Department of Defense

DIRECTIVE

NUMBER 3000.09
November 21, 2012

USD(P)

SUBJECT: Autonomy in Weapon Systems

References: See Enclosure 1

1. PURPOSE. This Directive:

a. Establishes DoD policy and assigns responsibilities for the development and use of autonomous and semi-autonomous functions in weapon systems, including manned and unmanned platforms.

b. Establishes guidelines designed to minimize the probability and consequences of failures in autonomous and semi-autonomous weapon systems that could lead to unintended engagements.

2. APPLICABILITY. This Directive:

a. Applies to:

(1) OSD, the Military Departments, the Office of the Chai
Staff and the Joint Staff (CJCS), the Combatant Command
of the Department of Defense, the Defense Agencie
organizational entities within the DoD (herei
Components").

ECRETARY OF THE NAVY
SHINGTON DC 20350-1000

June 5, 2015

F NAVAL OPERATIONS
NDANT OF THE MARINE CORPS

e and Robotics for Support Functions

(DON) is on the forefront of artificial intelligence (AI)
search. Recent operational examples of the DON's application of AI and
robotics technology include: Ghost Swimmer, an unmanned underwater vehicle that
mimics a Bluefin tuna; the X-47B Unmanned Combat Air System that can autonomously
land aboard an aircraft carrier; and the Swarmboat unmanned surface vehicle that can sync
with other unmanned vessels to swarm and interdict enemy vessels.

The private sector is investing heavily in AI and robotics automation for decision-
making and physical implementation tasks. The DON could benefit from considering how
to adapt recent private sector advances in fields such as machine learning, natural
processing, ontological engineering, and automated planning for

To accelerate the exploration of these
shall:

2. APPLICABILITY. This Directive:

a. Applies to:

(1) OSD, the Military Departments, t
Staff and the Joint Staff (CJCS), the Combata
of the Department of Defense, the Defense A
organizational entities within the DoD (herei
Components").

(2) The design, development, acquisition, testing, fielding, and employment of autonomous and semi-autonomous weapon systems, including guided munitions that can independently select and discriminate targets.

(3) The appli
semi-autonomous we

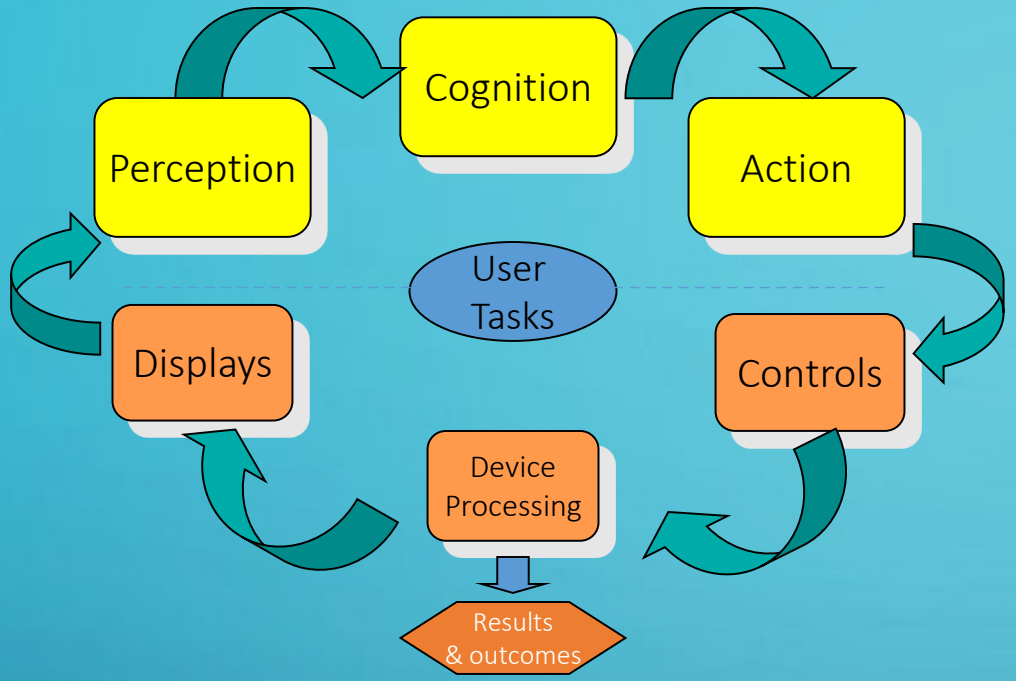
4. POLICY. It is DoD policy that:

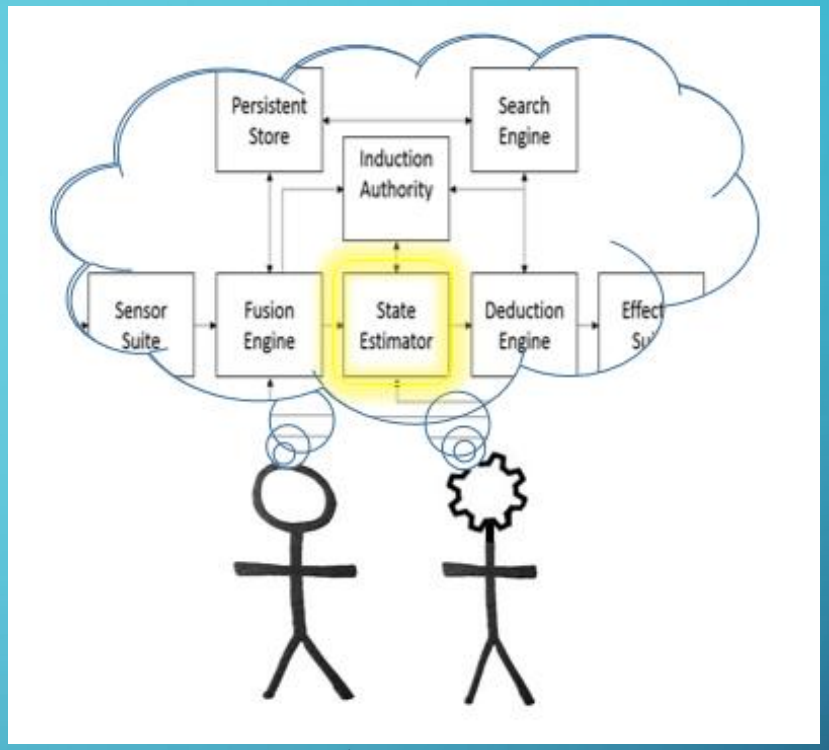
a. Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force.

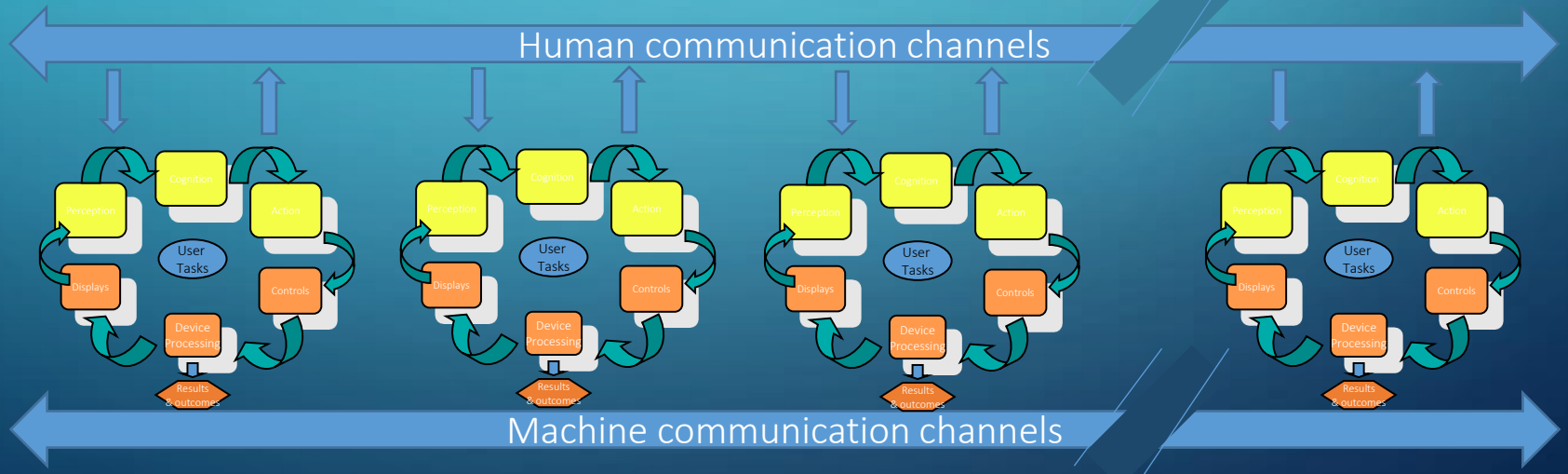
(2) Consistent with the potential consequences of an unintended engagement or loss of control of the system to unauthorized parties, physical hardware and software will be designed with appropriate:

(a) Safeties, anti-tamper mechanisms, and information assurance in accordance with DoD Directive 8500.01E (Reference (a)).

(b) Human-machine interfaces and controls.

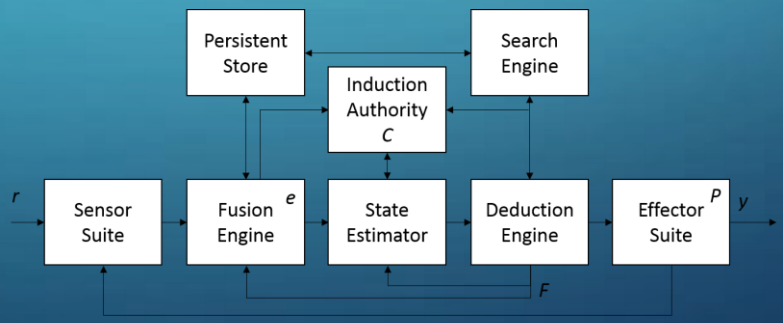

Cognition
Perception
Action
User Tasks
Displays
Controls
Device Processing
Results & outcomes


Persistent Store
Search Engine
Induction Authority
Sensor Suite
Fusion Engine
State Estimator
Deduction Engine

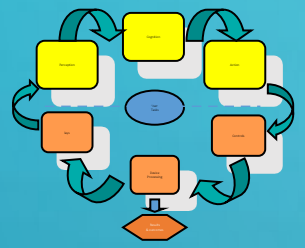

Human communication channels
Machine communication channels

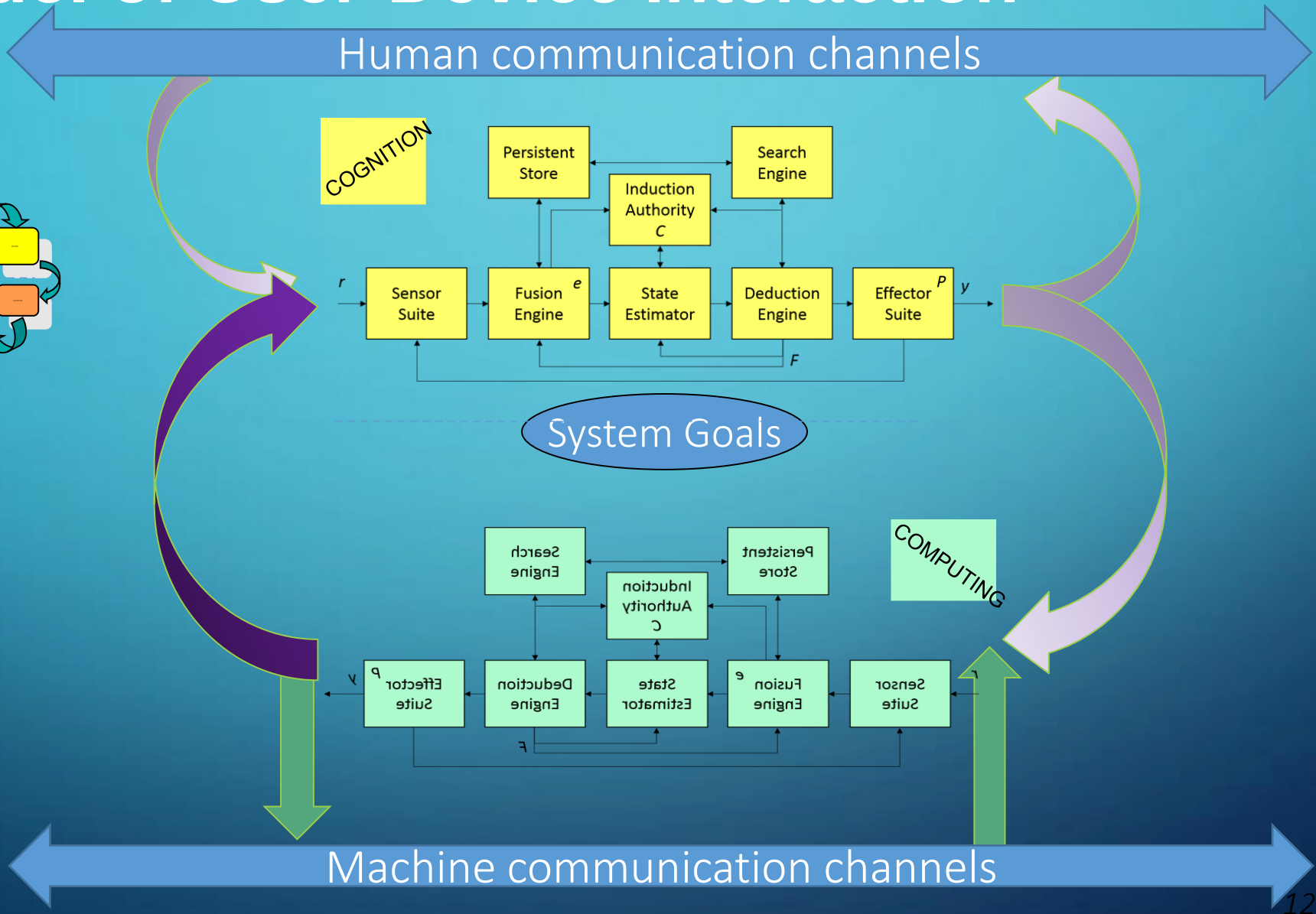
REQUIREMENT: DESIGN A CYBORG

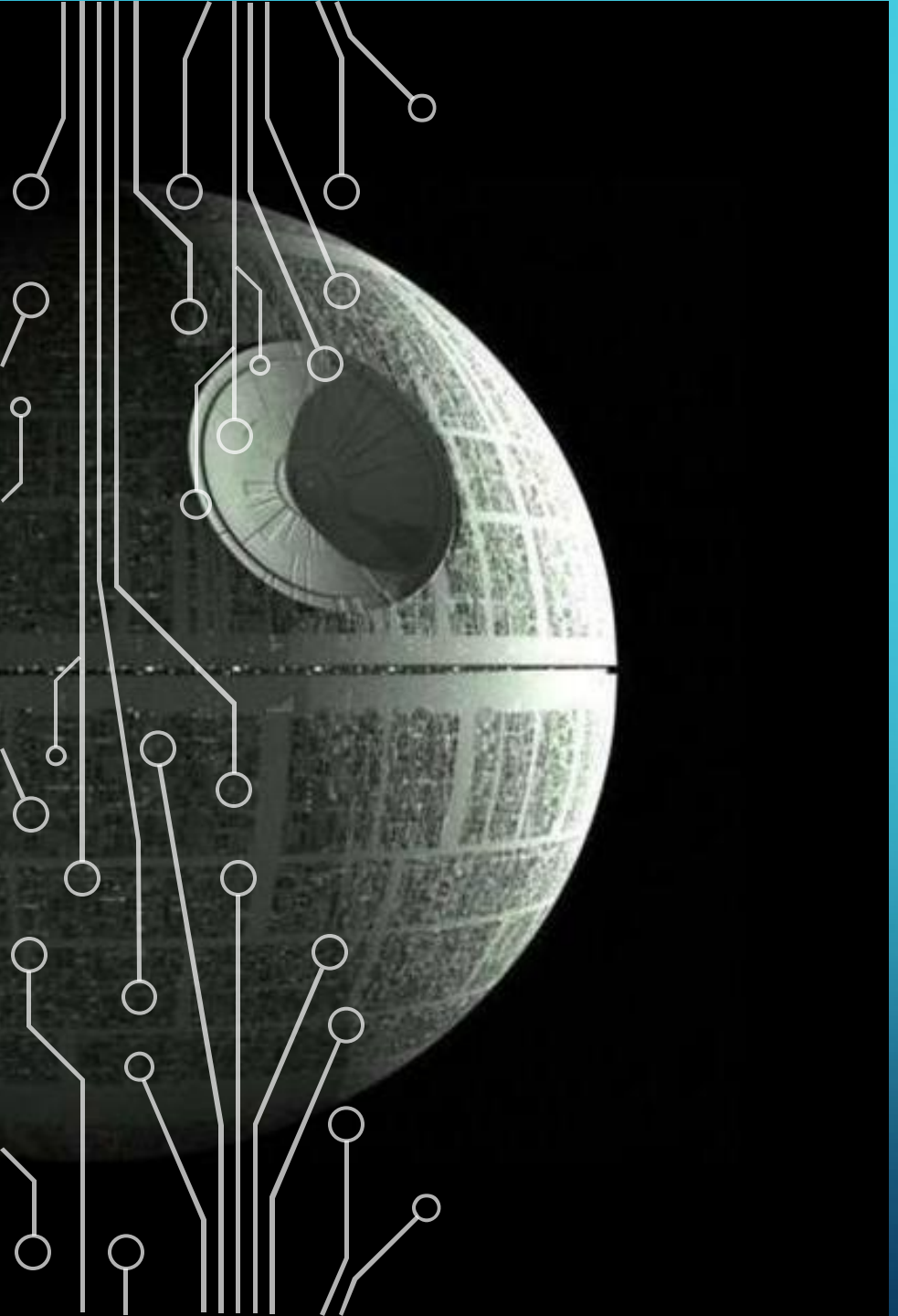
- Implied goal of Human-System Integration (HSI)
- More than the sum of its parts (hence × not +)
  - Emergent capabilities
  - A new entity



Model of User-Device Interaction







INTELLIGENT CONTROL THEORY CONFRONTS THE EFM PROBLEM

- *Intelligence* is required/demonstrated when the internal representation (model, simulation) has more parameters than observables
  - *Something* is required to identify and estimate the parameters
  - Missing parameters are not deducible from sense data
  - There is no single “correct” solution – this is currently the only viable answer
- Theory of simulation meets the need for a theory of internal representation
- Graph theory is the mathematical

ASSURING HUMAN CONTROL AUTHORITY

- Analysis shows EFM are unpredictable and nearly inevitable in complex systems (including SOA/SoS)
- We must design systems so human and machine components can work together as seen in both experience and policy
- We set out to identify the necessary, if not minimum necessary, and sufficient features of systems to meet the DODDIR requirement
- Solutions will arise from focusing on representing where the human – machine entity is, relative to its goal
  - Apportioning control processes to human and machine components to exploit human strengths to detect, diagnose and redress emergent failures
  - Assure that proposed system design conforms to Intelligent Control Theory process architecture