

*National Academy of Sciences workshop for Air Force Research Laboratories
Human-AI Teaming Through Warfighter-Centered Designs, July 28, 2021*

Human-Centered AI: Reliable, Safe & Trustworthy

Ben Shneiderman @benbendc

Founding Director (1983-2000), Human-Computer Interaction Lab
Professor, Department of Computer Science

Member, National Academy of Engineering



Photo: BK Adams

What is Human-Centered AI?



What is Human-Centered AI?



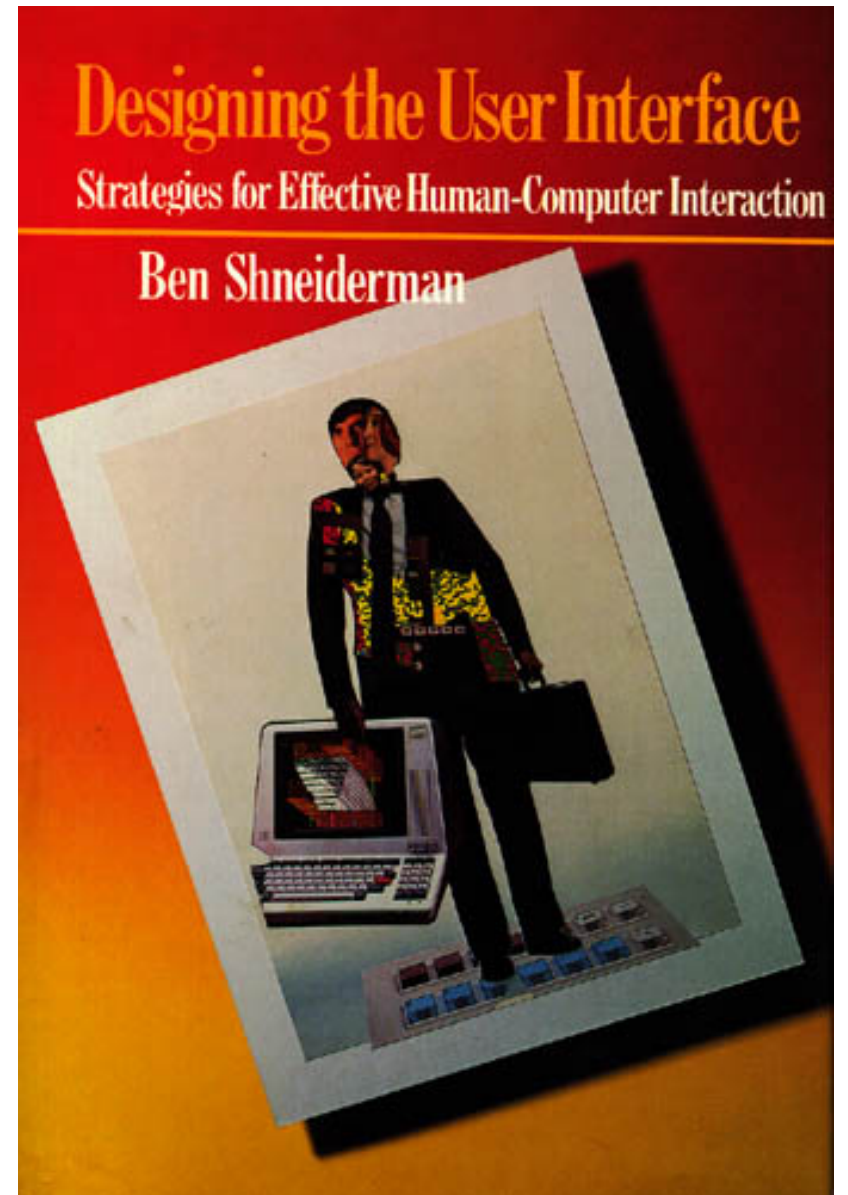
Amplify, Augment, Empower & Enhance People

HCAI Framework



Designing the User Interface

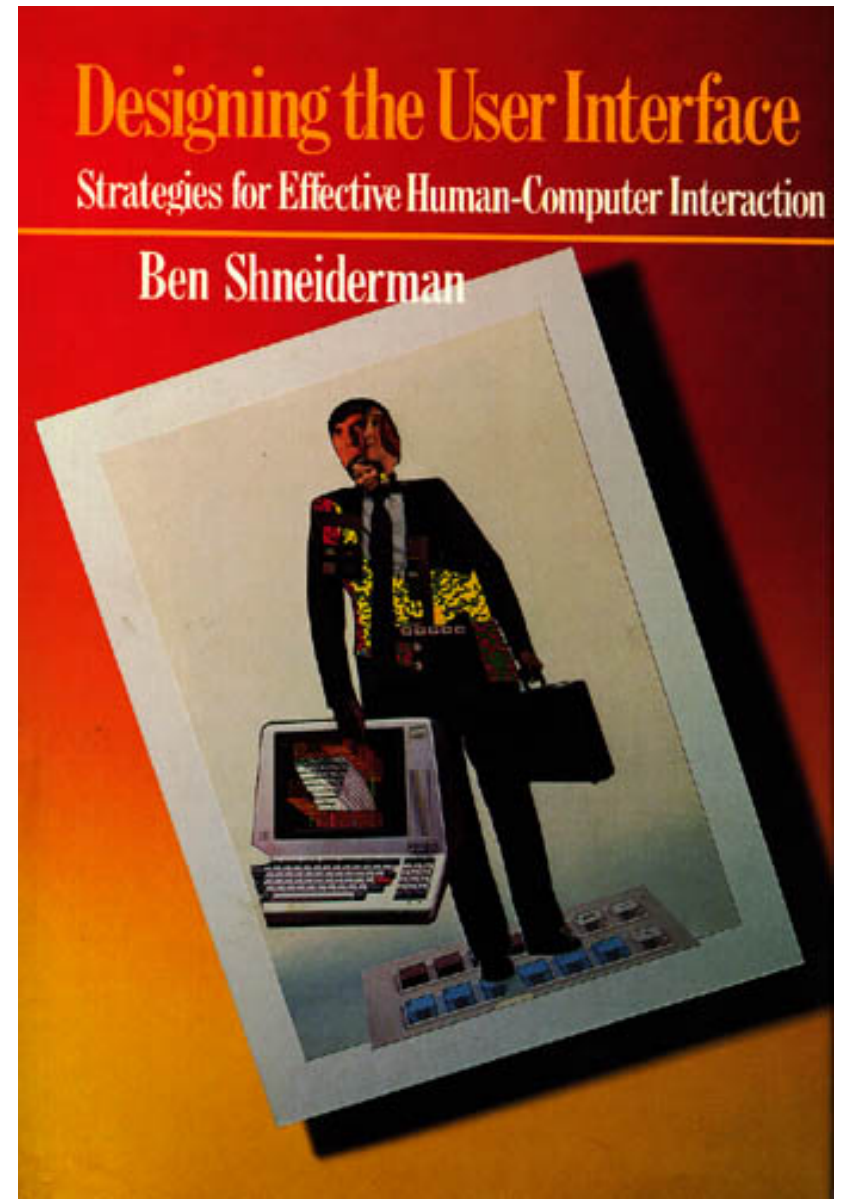
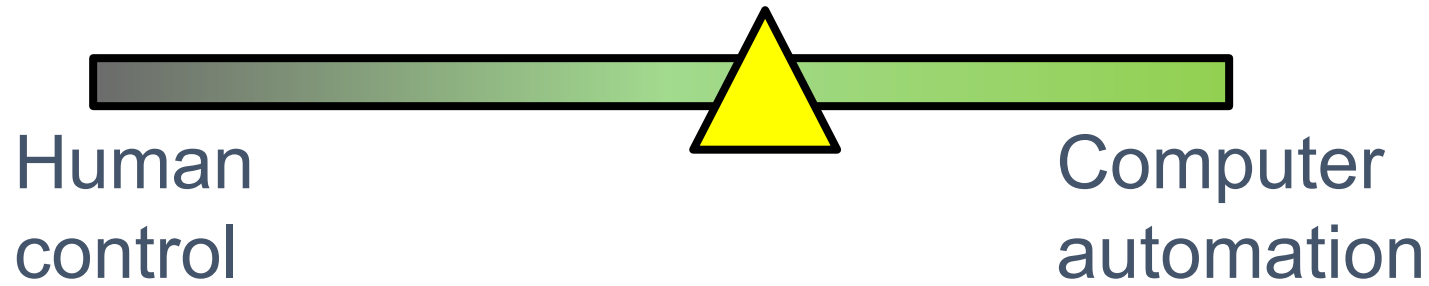
Balancing automation & human control



First Edition: 1986

Designing the User Interface

Balancing automation & human control



First Edition: 1986

Designing the User Interface

Ensuring human control
while increasing automation



Sixth Edition: 2016

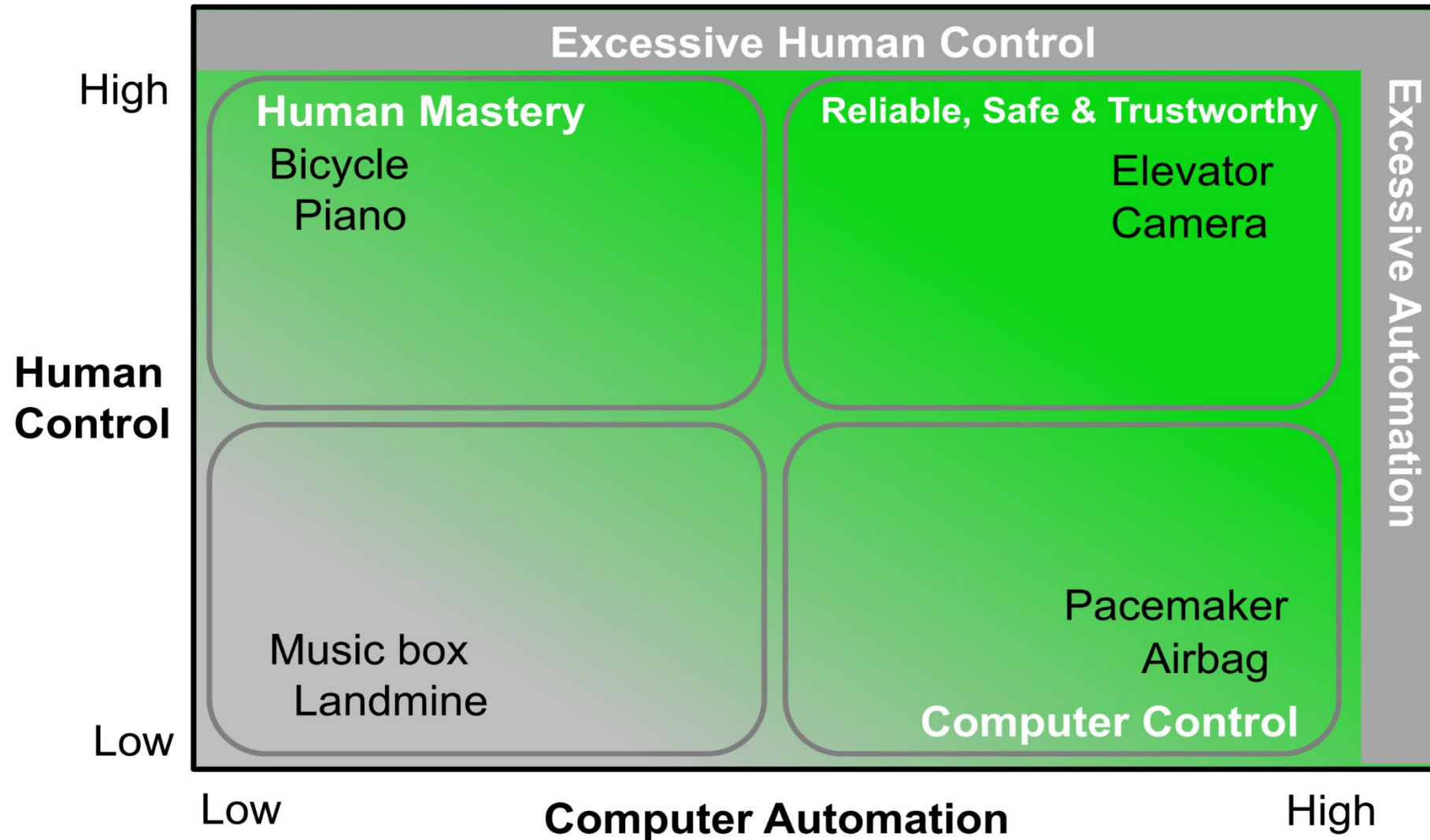
Designing the User Interface

Ensuring human control
while increasing automation



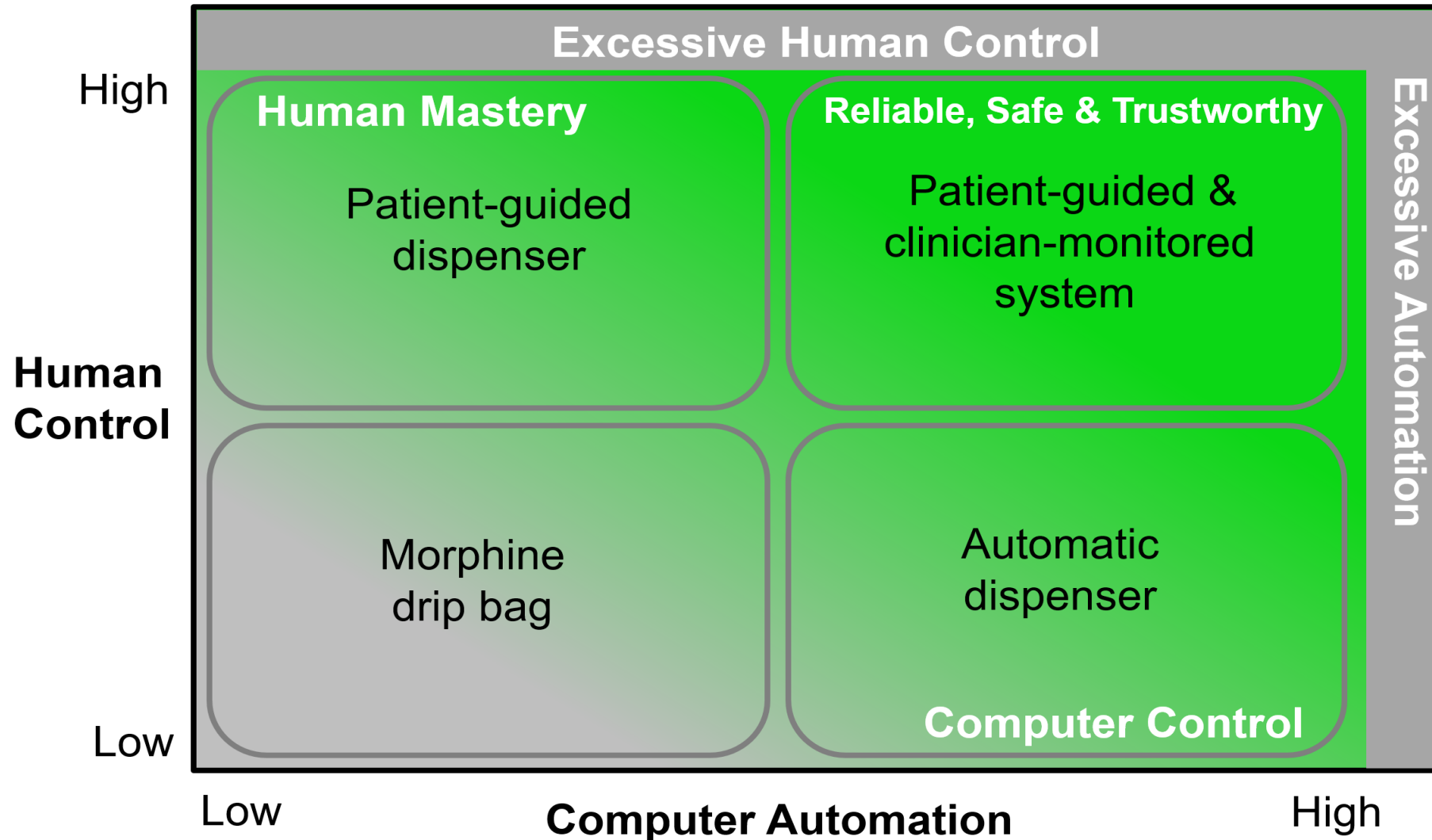
Sixth Edition: 2016

HCAI Framework



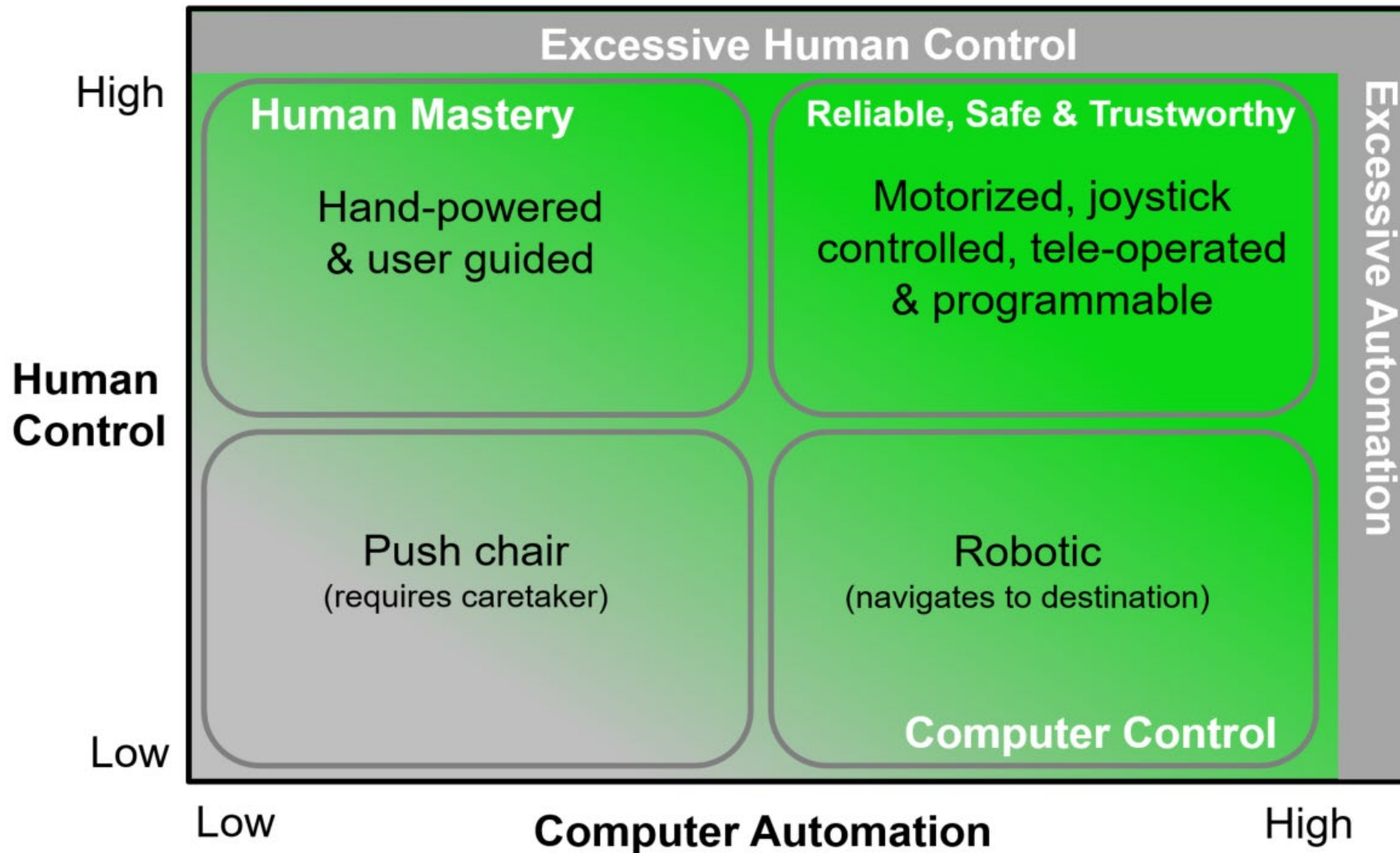
HCAI Framework

Pain Control Designs



HCAI Framework

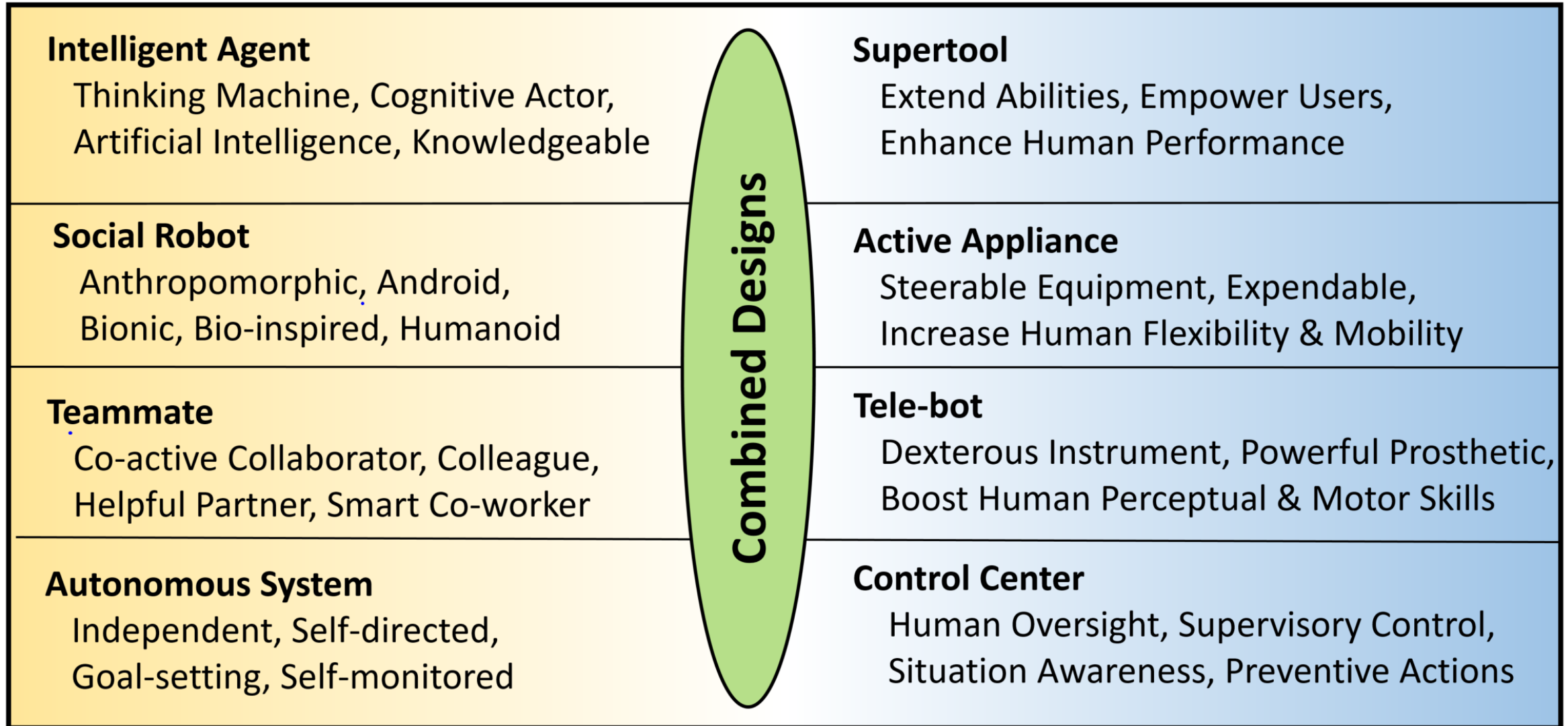
Wheelchair Designs



Design Metaphors



Design Metaphors



“Robots are simply not people”: Margaret Boden

“Humans, not robots, are responsible agents”: Joanna Bryson

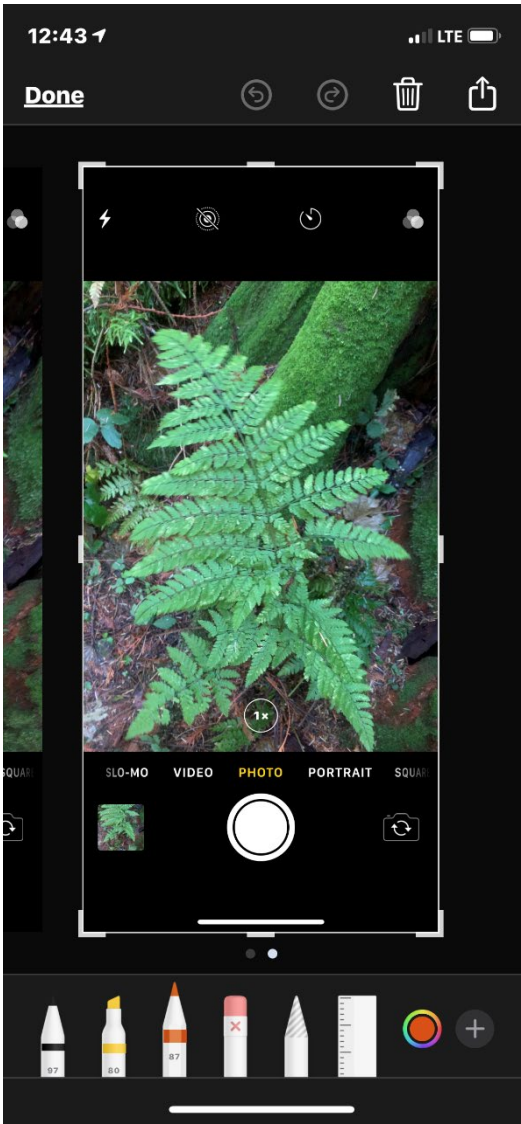
- **Responsibility:** Only humans are liable, legally and morally
- **Distinctive capabilities of computers:** sophisticated algorithms, huge databases, superhuman sensors, information abundant displays, & powerful effectors
- **Human creativity:** passion, empathy, humility & intuition

DoD Directive 3000.09: Autonomy in Weapon Systems

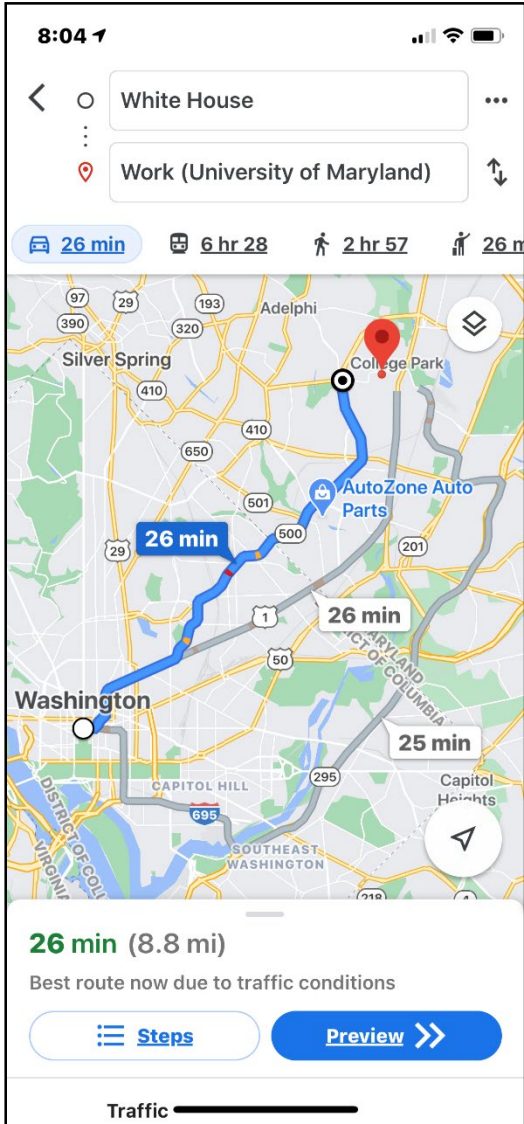
- Autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force.
- Designed with appropriate ... human-machine interfaces and controls
- Design autonomous and semi-autonomous weapon systems in such a manner as to minimize the probability and consequences of failures that could lead to unintended engagements or to loss of control of the system.
- Readily understandable to trained operators, provide traceable feedback on system status, and provide clear procedures for trained operators to activate and deactivate system functions

Supertools

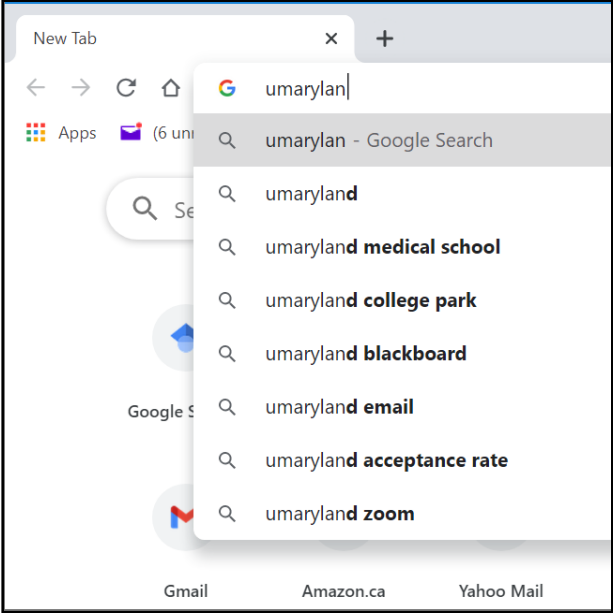
Digital Camera Controls



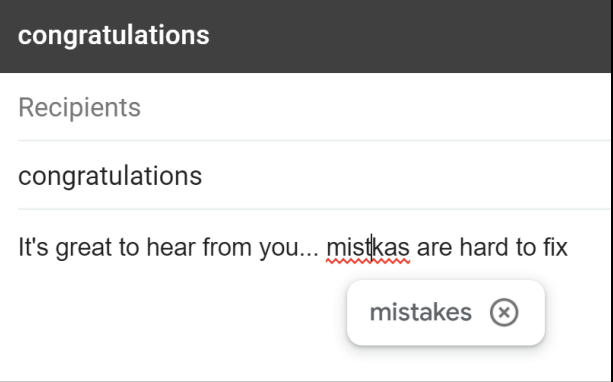
Navigation Choices



Texting/Search Autocompletion



Spelling correction



Supertool: Bloomberg Terminal



Active Appliances

Coffee maker, Rice cooker, Blender

Dishwasher, Clothes Washer/Dryer



Cuisinart Grind & Brew Coffee Maker



Panasonic Rice Cooker



Nutri Ninja Blender



Miele Dishwasher

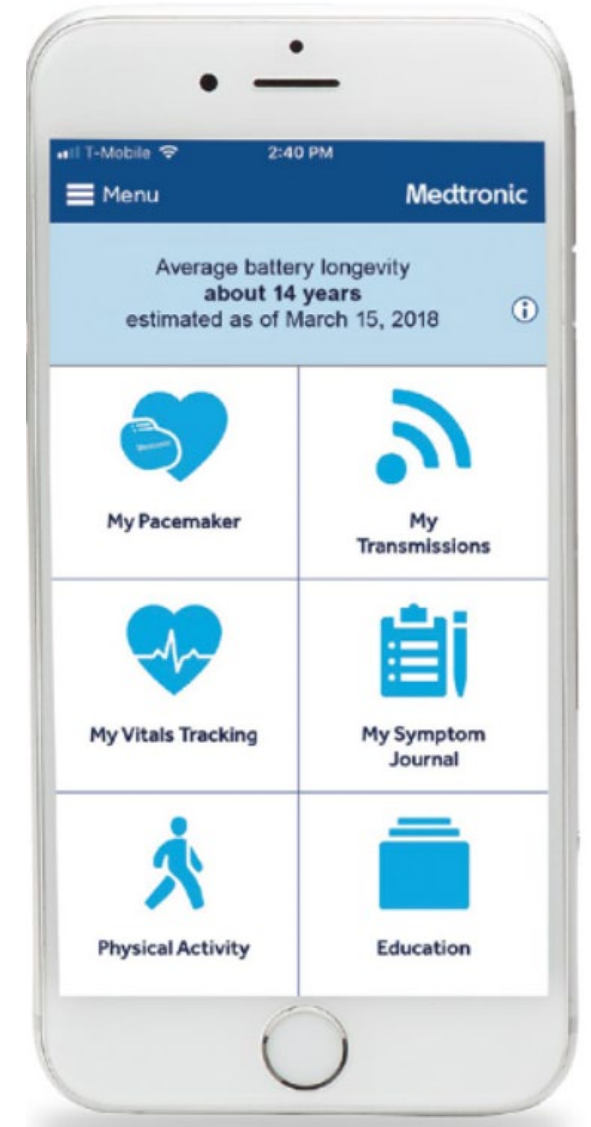
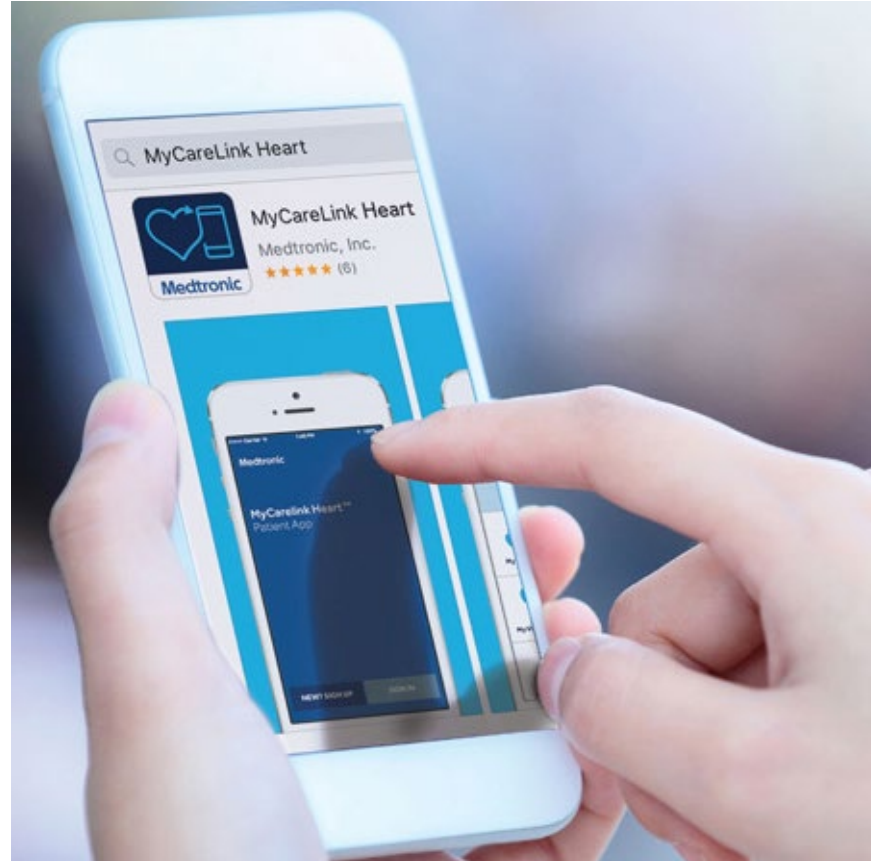
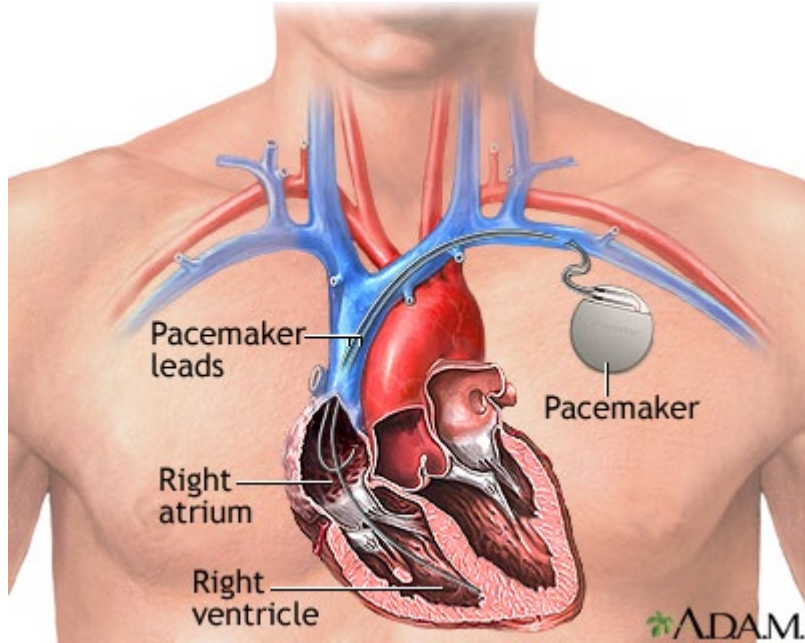


General Electric Washer



General Electric Dryer

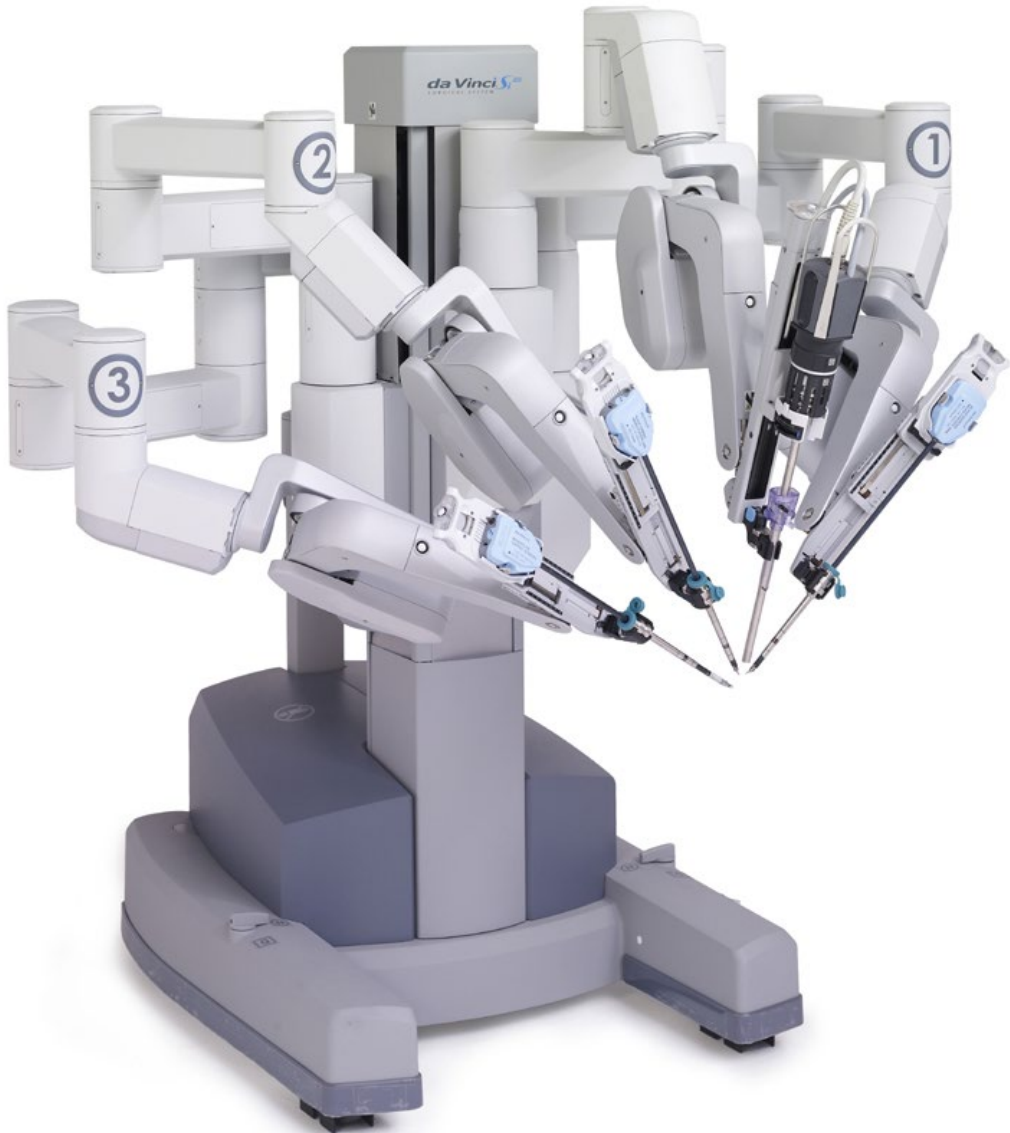
Active Appliance: Implanted Cardiac Pacemakers



Tele-Bot: NASA Mars Rovers



Tele-Bot: Da Vinci Surgical System



“Robots don’t perform surgery. Your surgeon performs surgery with da Vinci by using instruments that he or she guides via a console.”

<https://www.davincisurgery.com/>

Control Center: Hospital



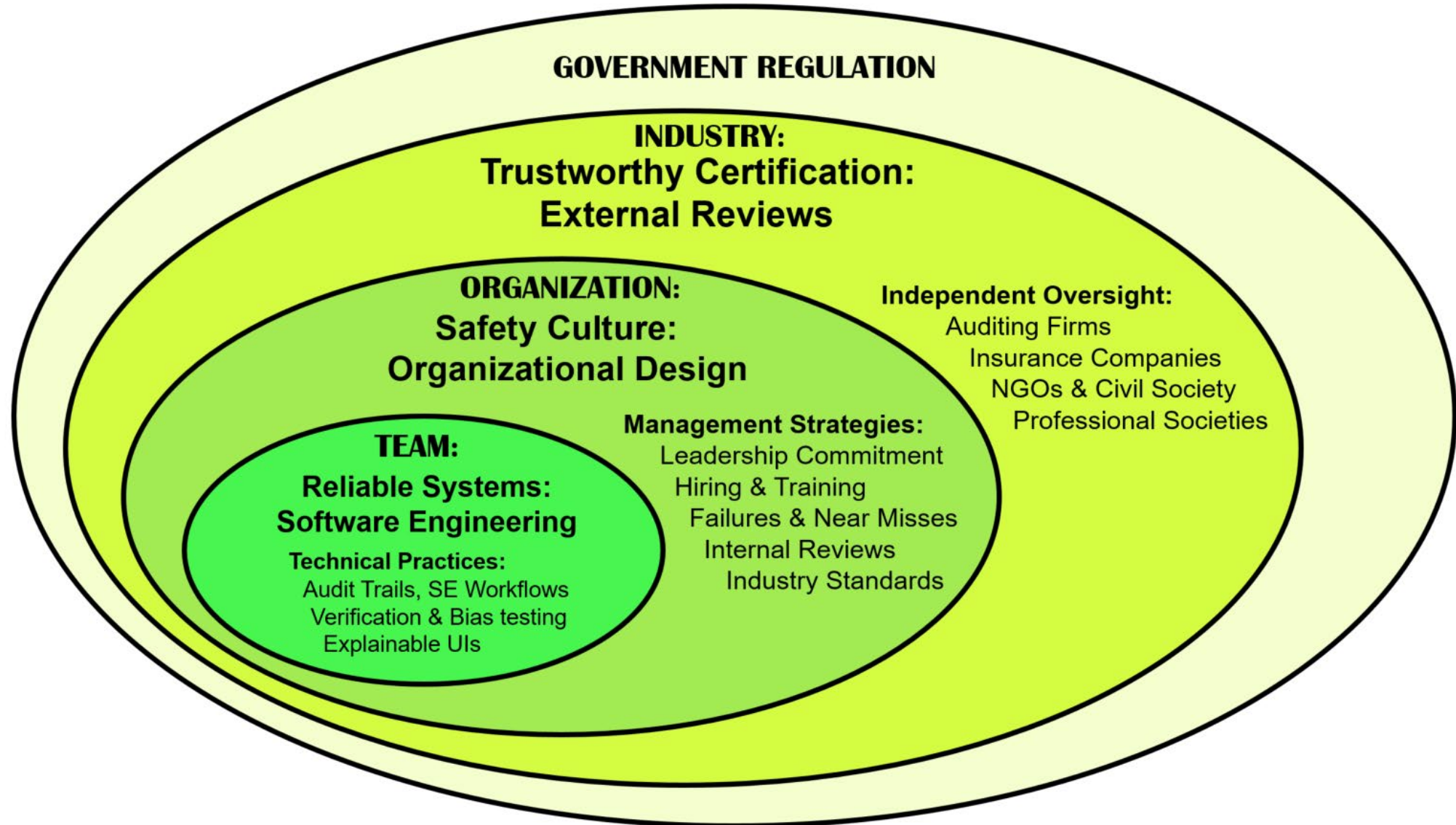
Control Center: Counter Terrorism



Governance Structures



Governance Structures for Human-Centered AI

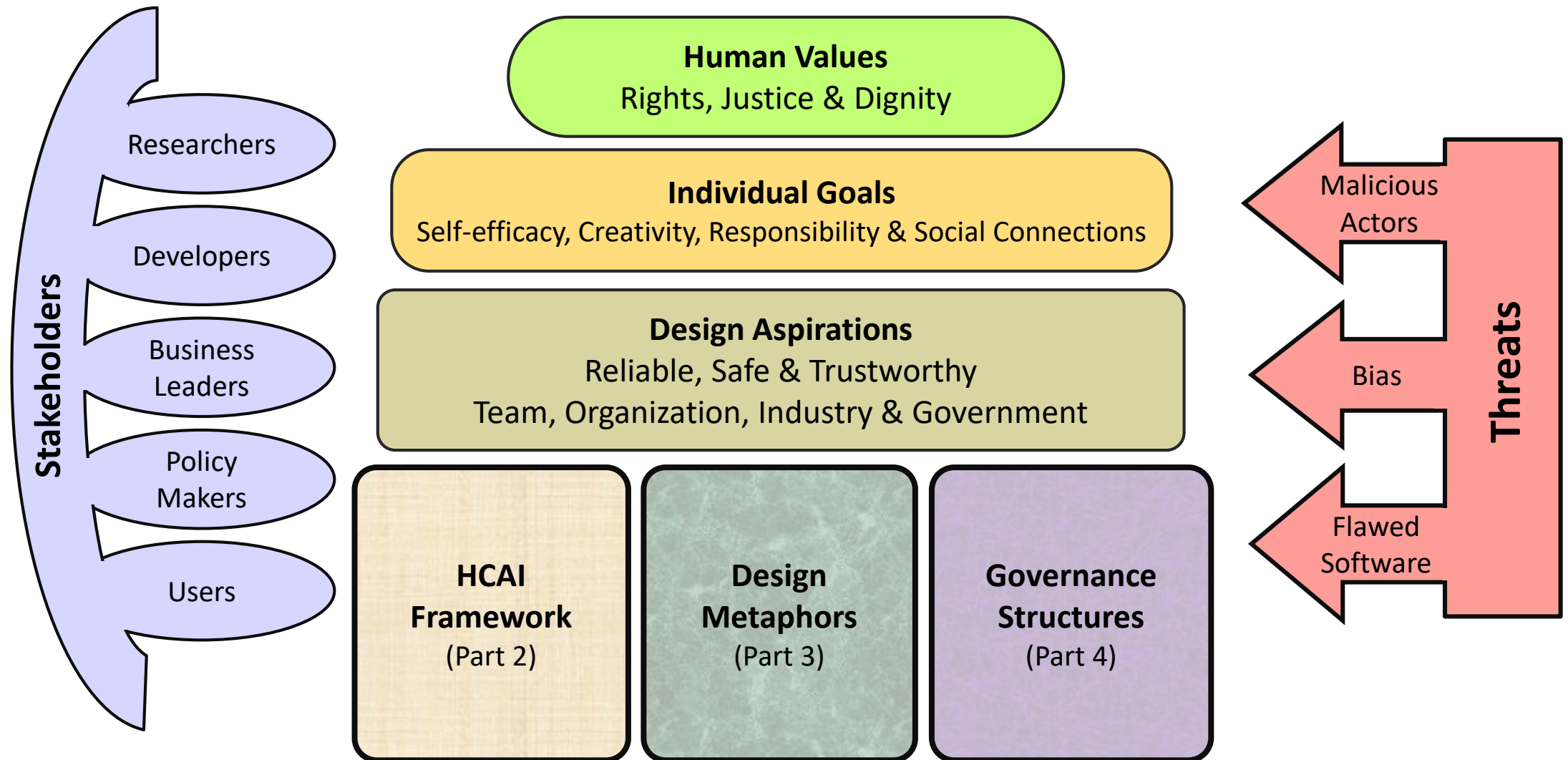




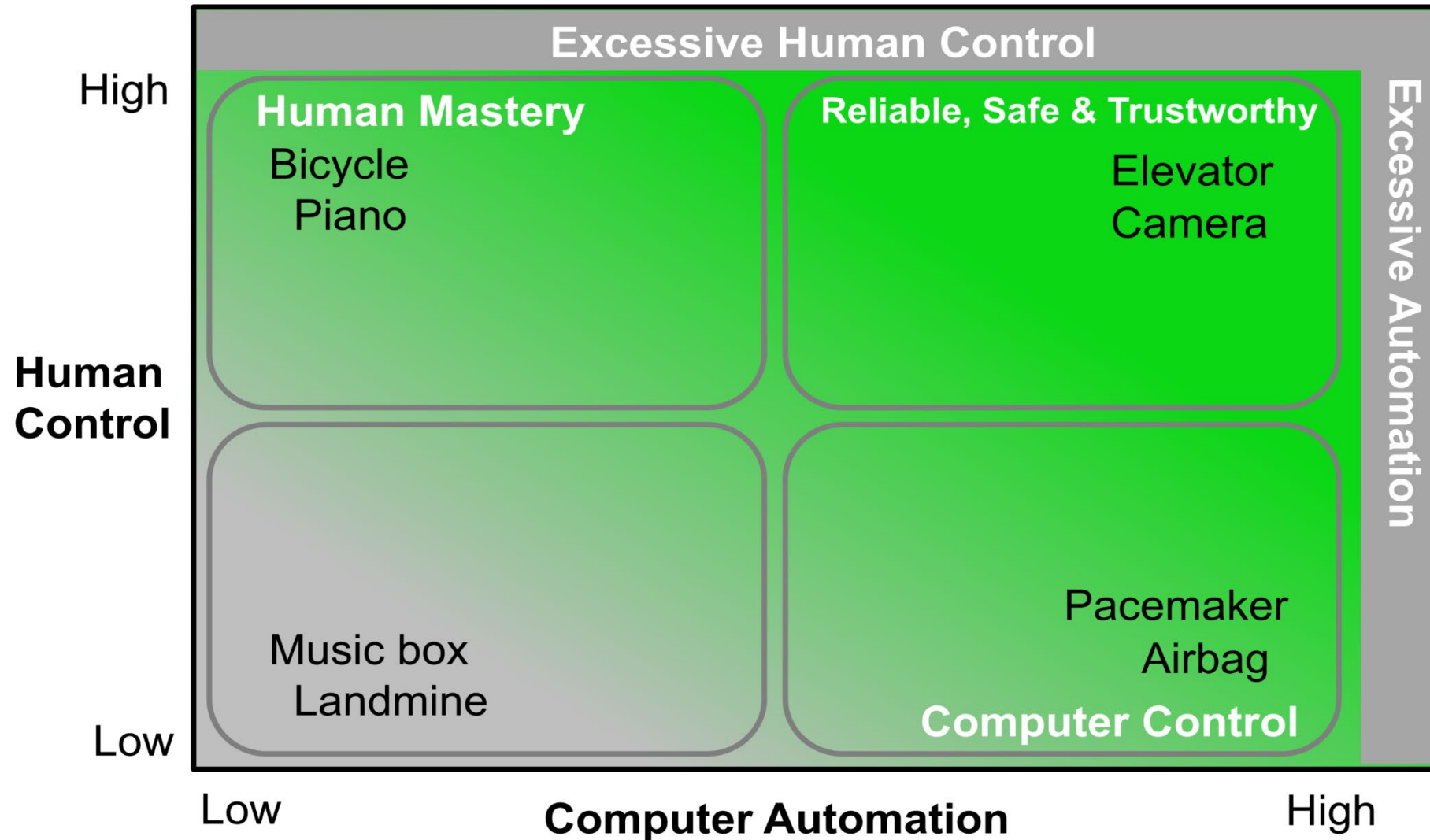
Summary



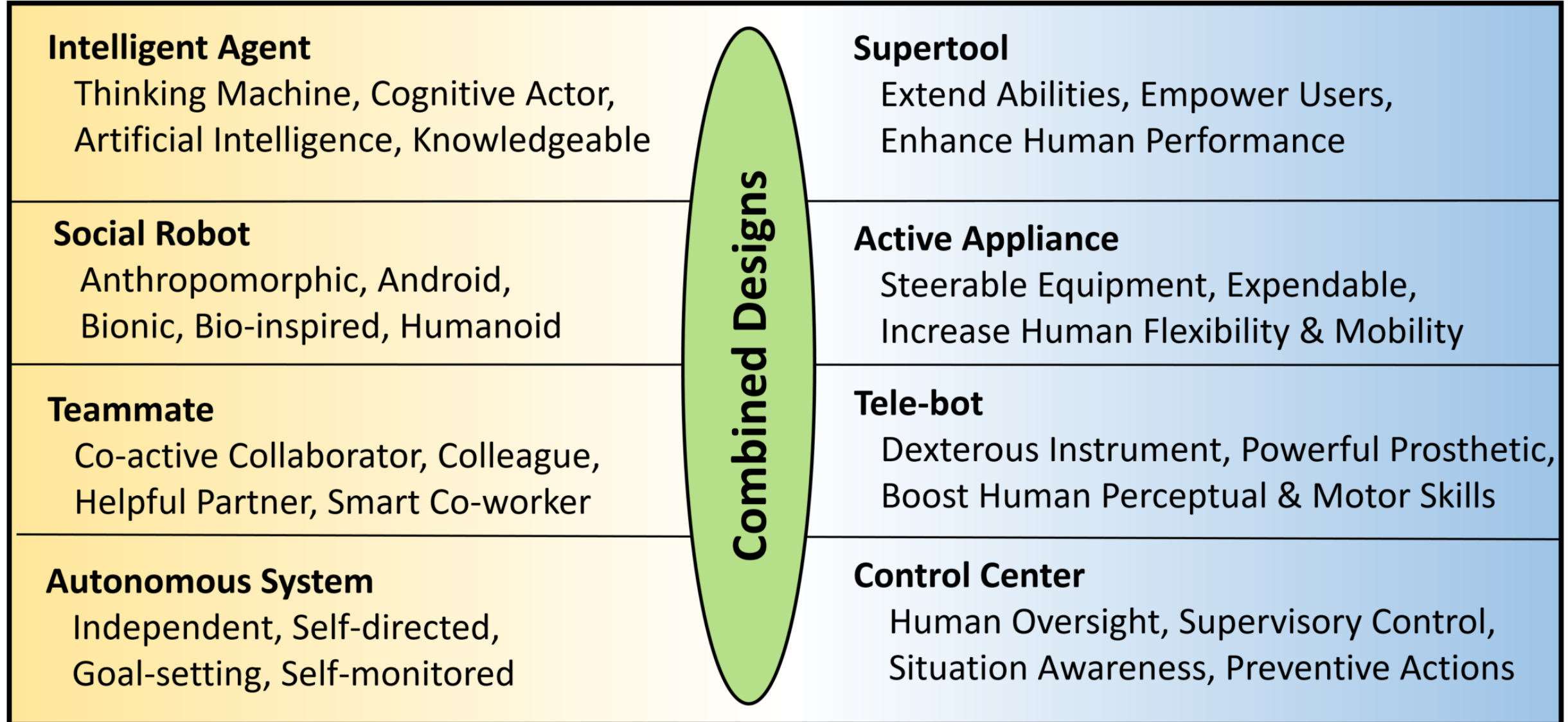
Human-Centered AI



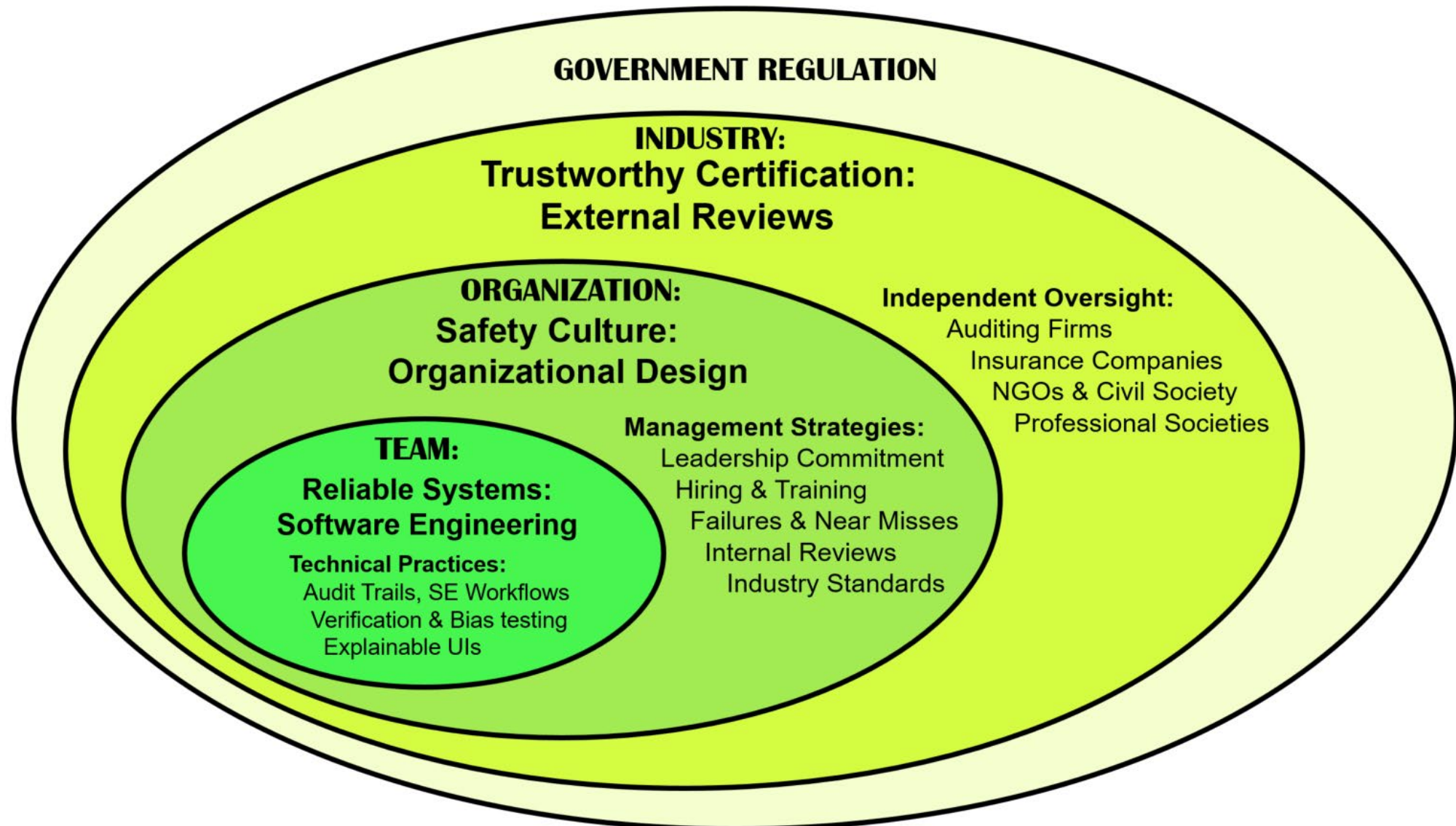
HCAI Framework



Design Metaphors



Governance Structures for Human-Centered AI



Human-Centered Artificial Intelligence: Reliable, safe & trustworthy, *International Journal of Human-Computer Interaction* 36, 6 (March 2020). <https://doi.org/10.1080/10447318.2020.1741118>

Design lessons from AI's two grand goals: Human emulation and useful applications, *IEEE Transactions on Technology & Society* 1, 2 (June 2020). <https://ieeexplore.ieee.org/document/9088114>

Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy Human-Centered AI systems, *ACM Trans. on Interactive Intelligent Systems* 10, 4 (Oct 2020). <https://dl.acm.org/doi/10.1145/3419764>

Human-Centered Artificial Intelligence: Three fresh ideas, *AIS Trans. on Human-Computer Interaction* 12, 3 (Oct 2020). <https://aisel.aisnet.org/thci/vol12/iss3/1/>

Human-Centered AI, *NAS ISSUES* 37, 2 (Winter 2021). <https://issues.org/human-centered-ai/>

Summary & resources: <https://hcil.umd.edu/human-centered-ai/>

The Future is Human-Centered

Google Group

<https://groups.google.com/g/human-centered-ai>

Twitter Account

@HumanCenteredAI

Website

<https://hcai.site>

[illegible]

Human-AI Teaming for Peacemakers/Peackeepers

- Negotiation, conflict resolution, discussion, deliberation
- Collaboration, partnership, diplomacy, cooperation

Yet, the National Security Commission on AI wrote

- **“to defend and compete in the coming era of AI-accelerated competition and conflict...to win the AI era”**
(35 instances of win)