

Swansea University, February 16, 2022

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





Interdisciplinary research community

- Computer Science & Info Studies
- Psych, Socio, Educ, Jour & MITH

hcil.umd.edu
vimeo.com/72440805

Designing the User Interface

Design Theories

Direct manipulation

Menus, speech, search

Social Media

Information Visualization

www.cs.umd.edu/hcil/DTUI6



Sixth Edition: 2016

What is Human-Centered AI?



What is Human-Centered AI?



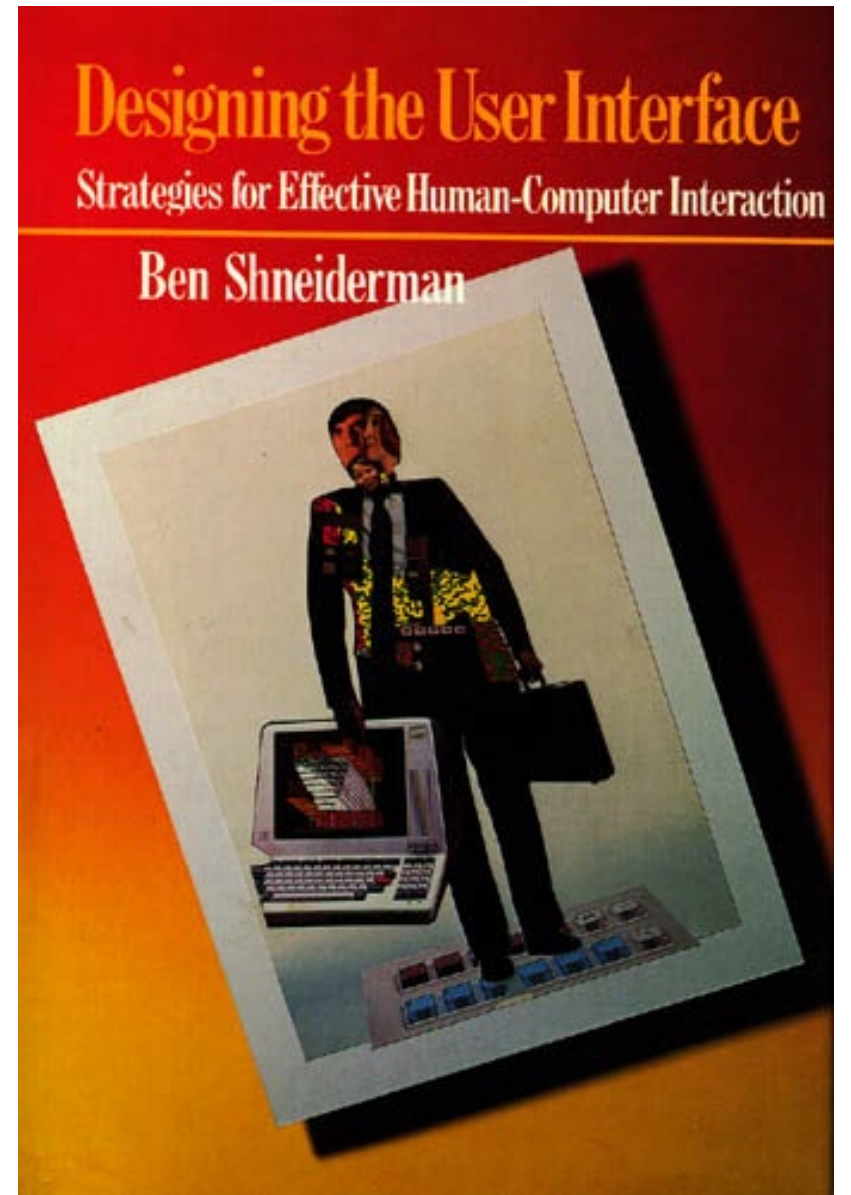
Amplify, Augment, Empower & Enhance People

HCAI Framework



Designing the User Interface

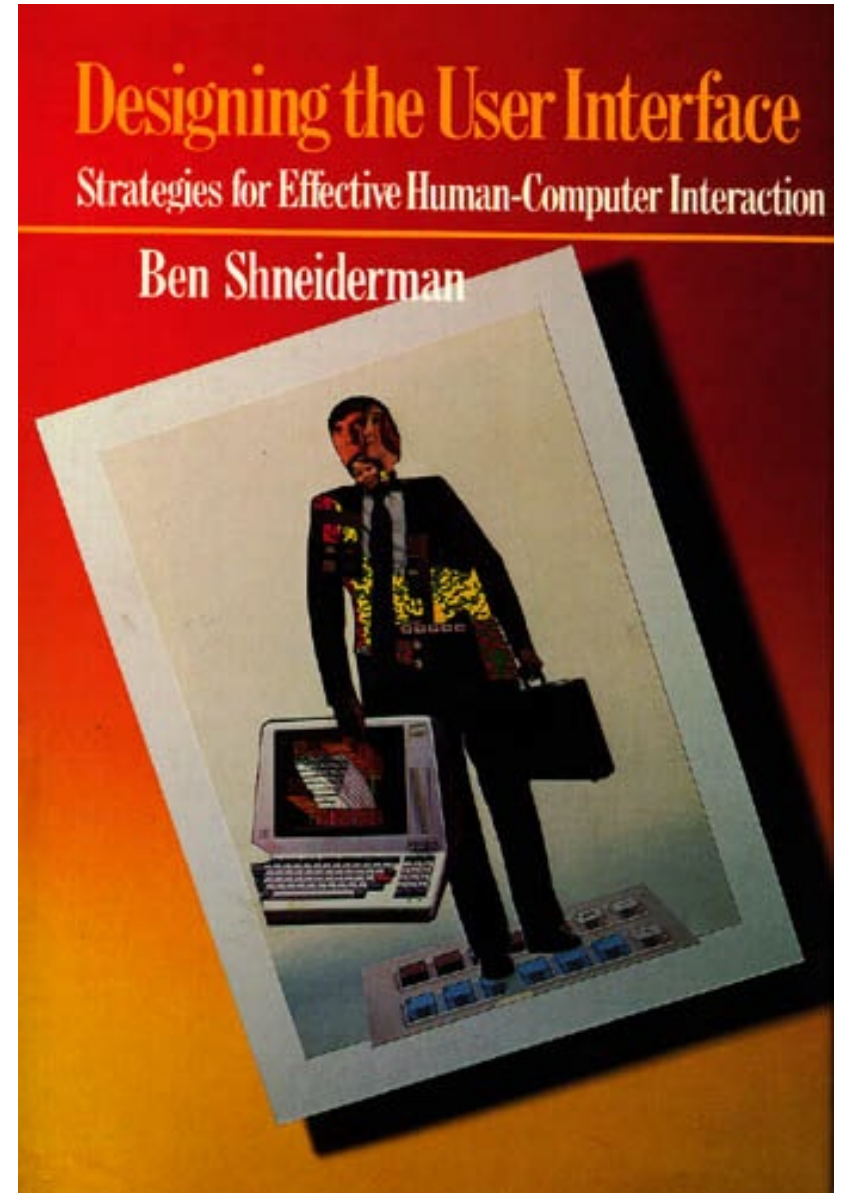
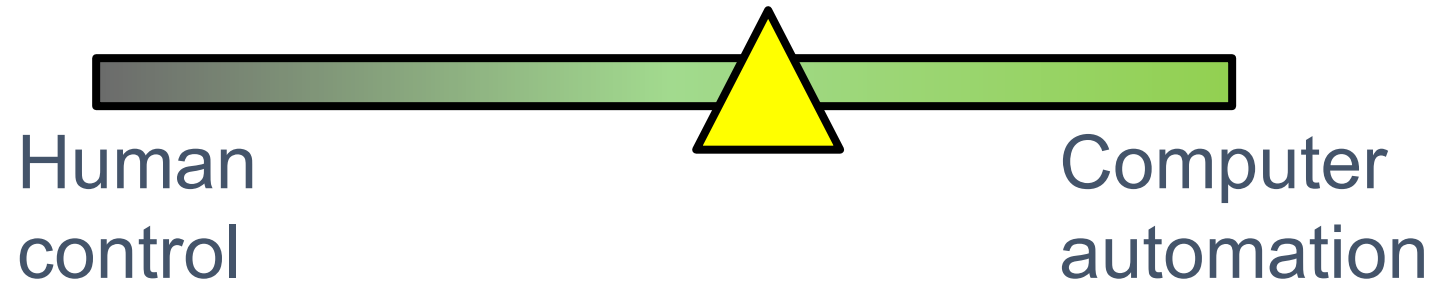
Balancing automation & human control



First Edition: 1986

Designing the User Interface

Balancing automation & human control



First Edition: 1986

Designing the User Interface

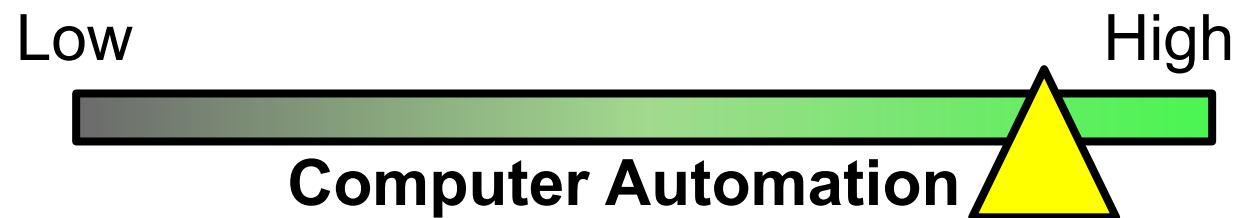
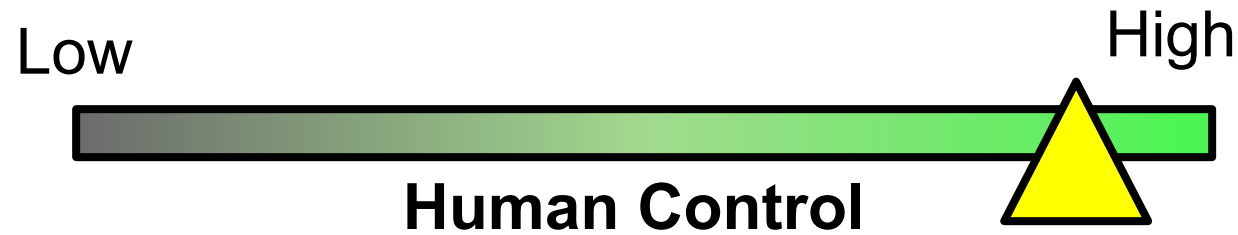
Ensuring human control
while increasing automation



Sixth Edition: 2016

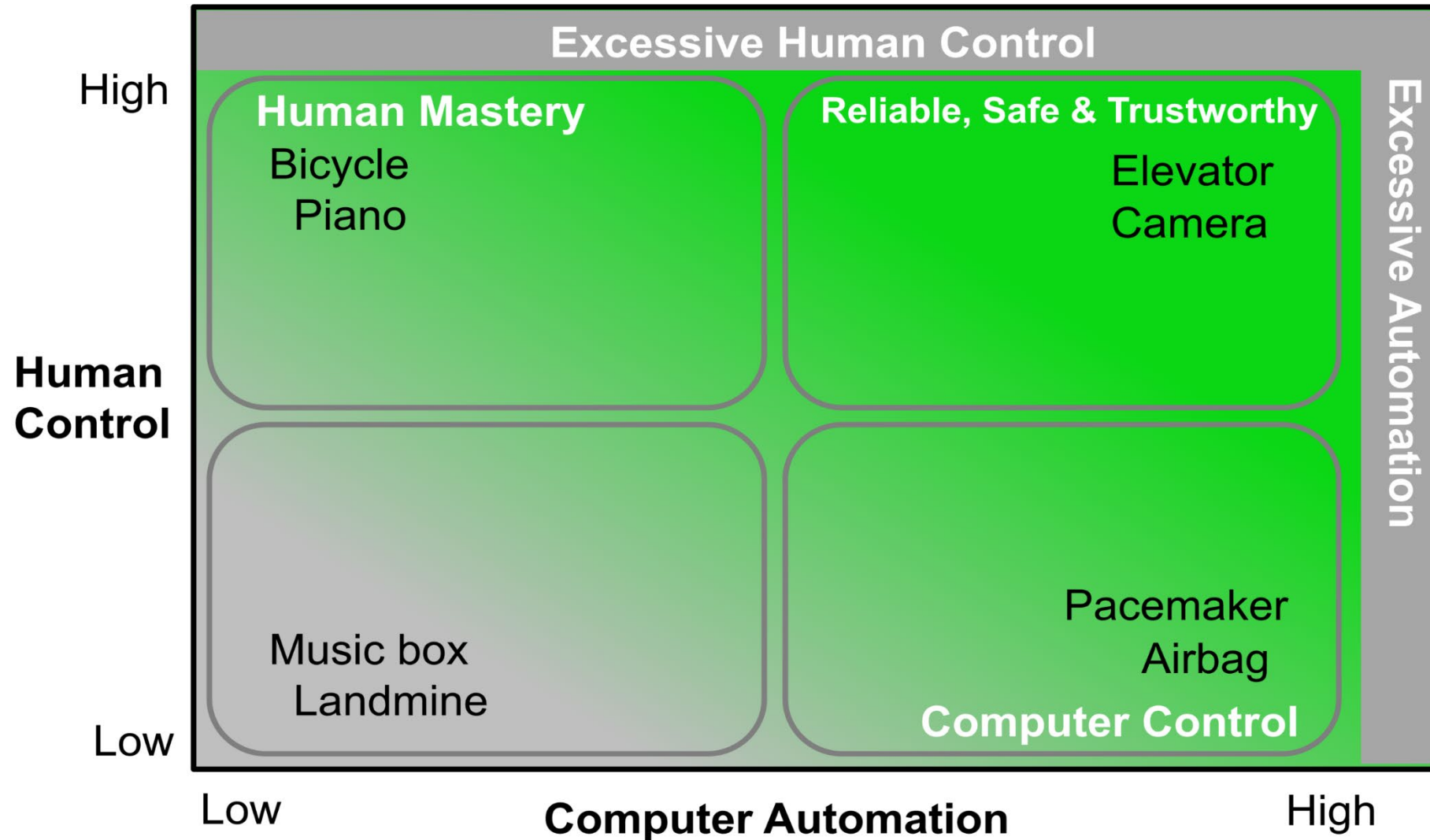
Designing the User Interface

Ensuring human control
while increasing automation



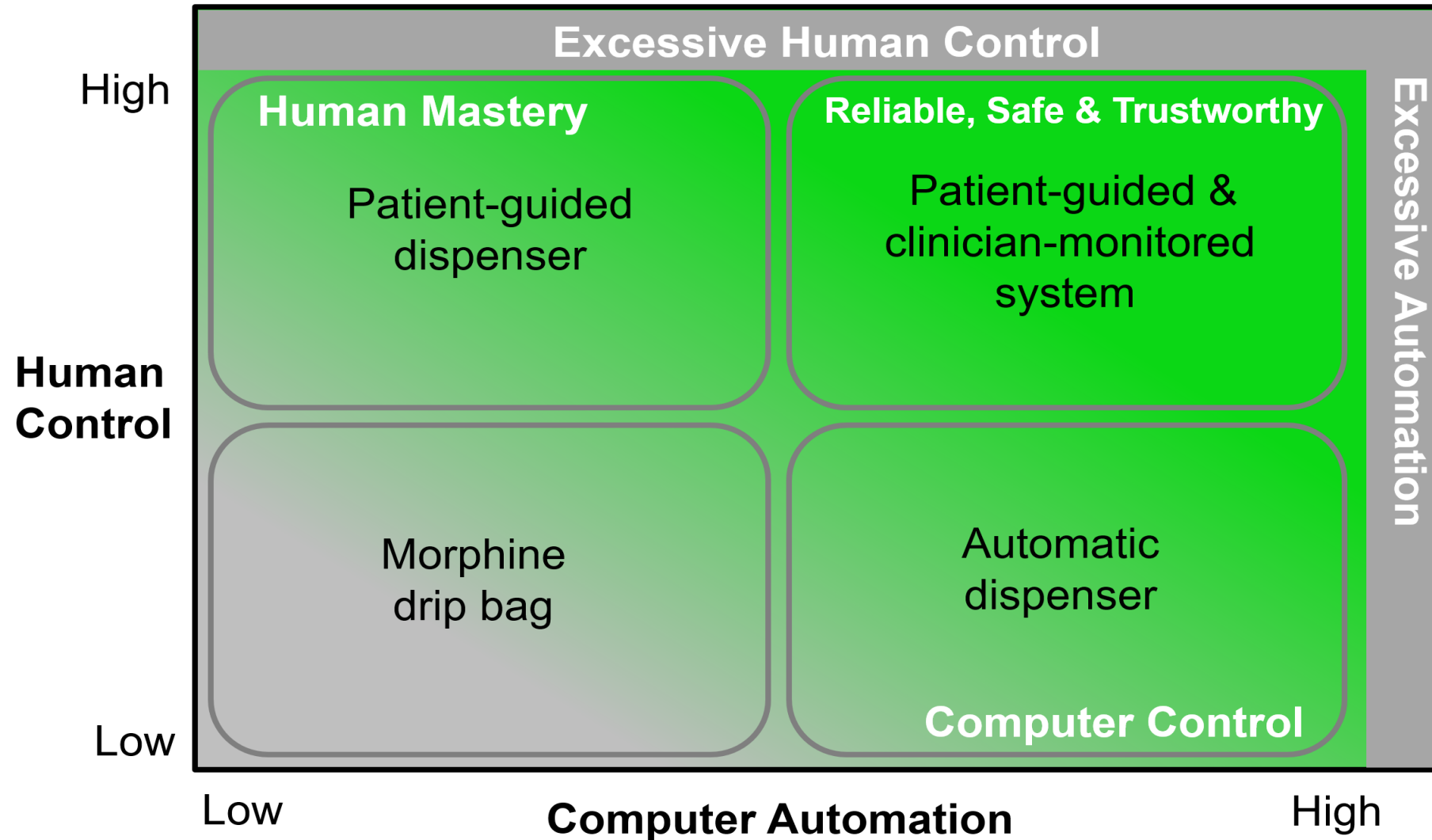
Sixth Edition: 2016

HCAI Framework



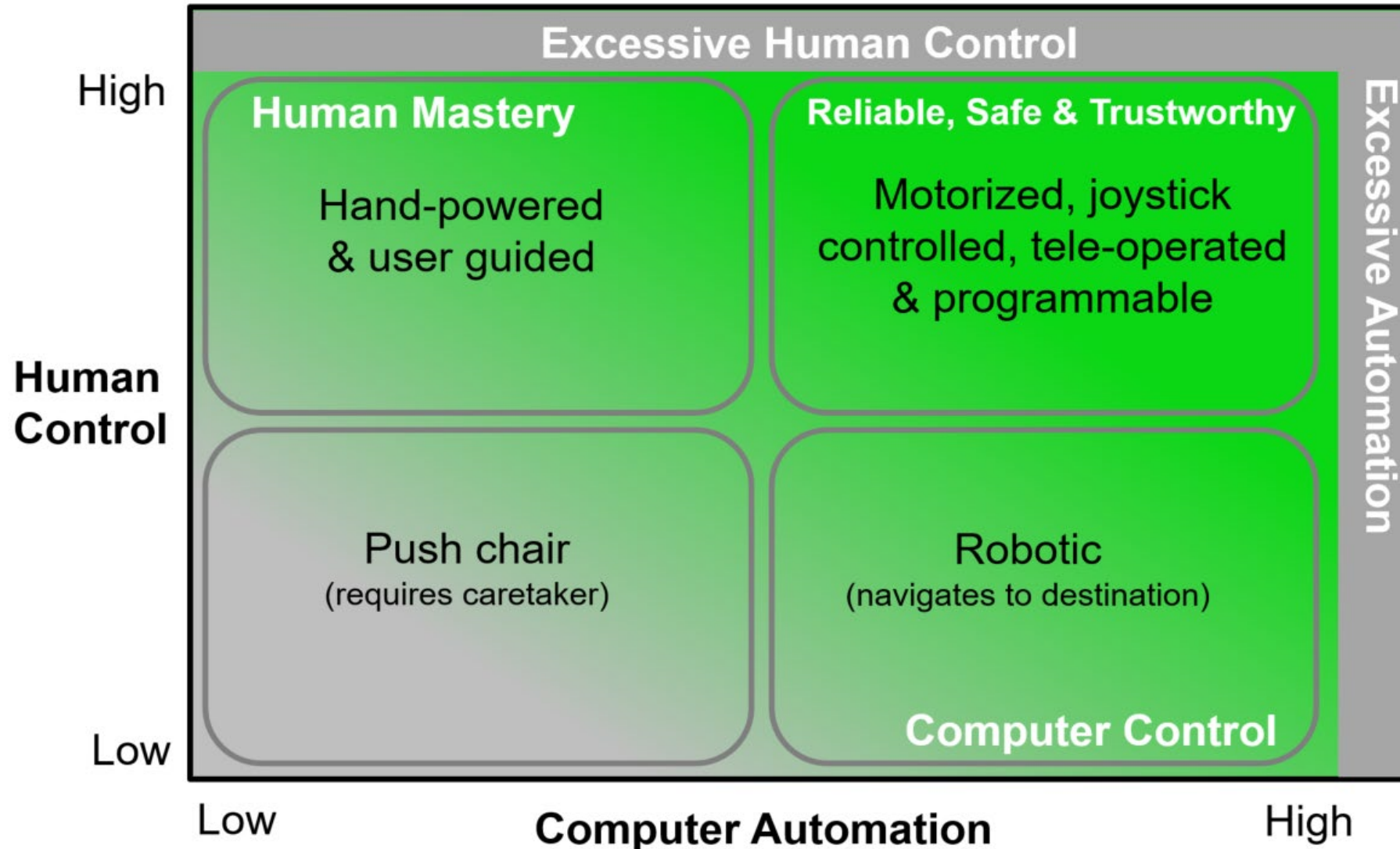
HCAI Framework

Pain Control Designs



HCAI Framework

Wheelchair Designs



Design Metaphors

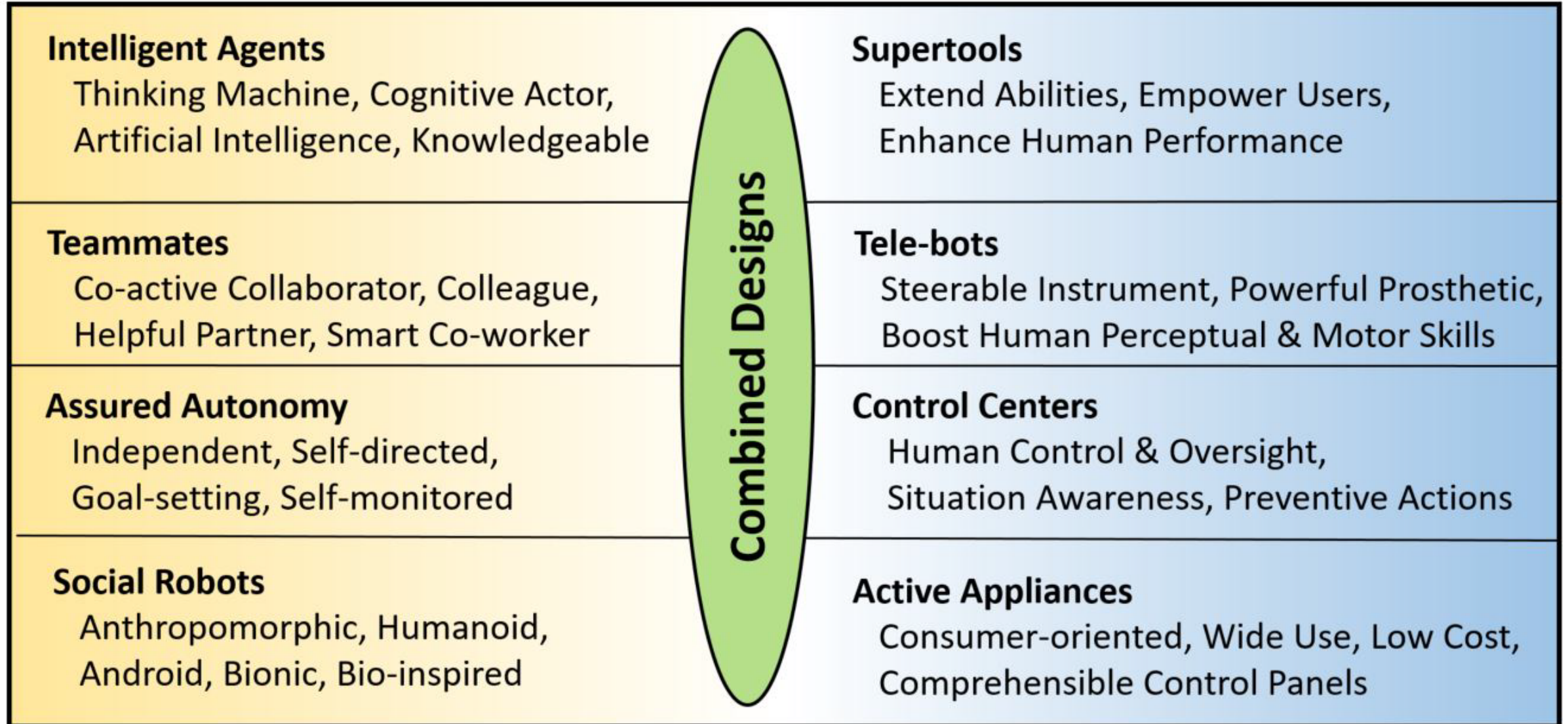


“Robots are simply not people”: Margaret Boden

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

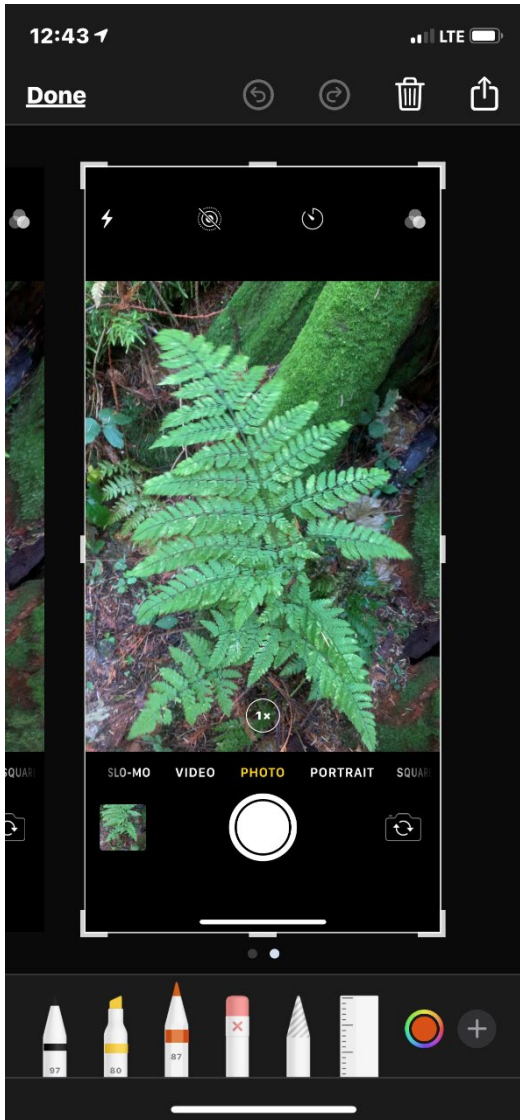
- **Responsibility:** Only humans are liable, legally and morally
- **Distinctive capabilities of computers:** sophisticated algorithms, huge databases, superhuman sensors, information abundant displays, & powerful effectors
- **Human creativity:** passion, empathy, humility & intuition

Design Metaphors

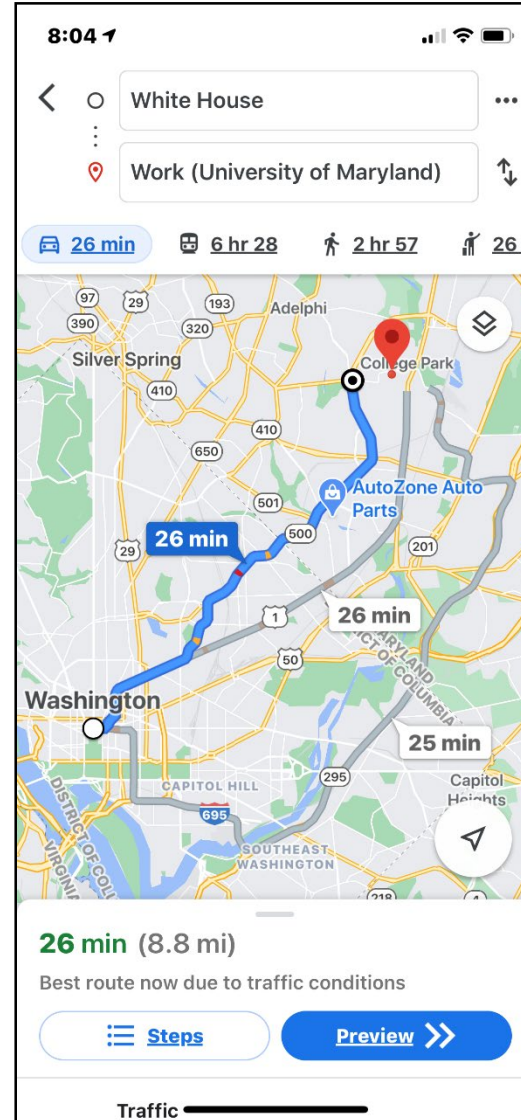


Supertools

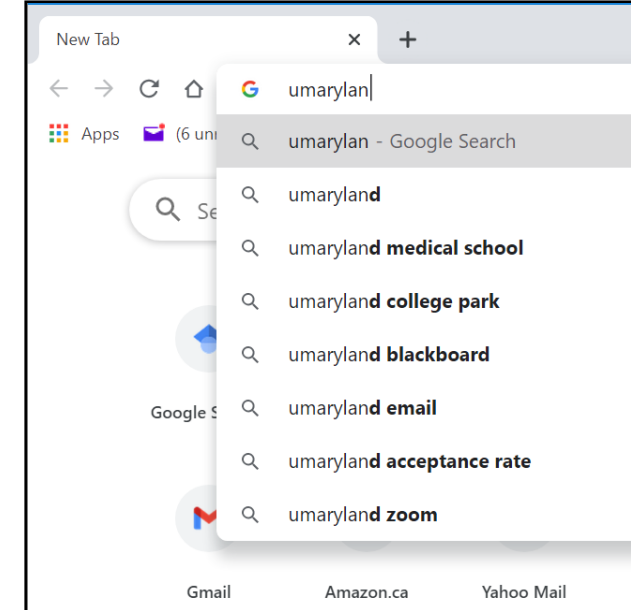
Digital Camera Controls



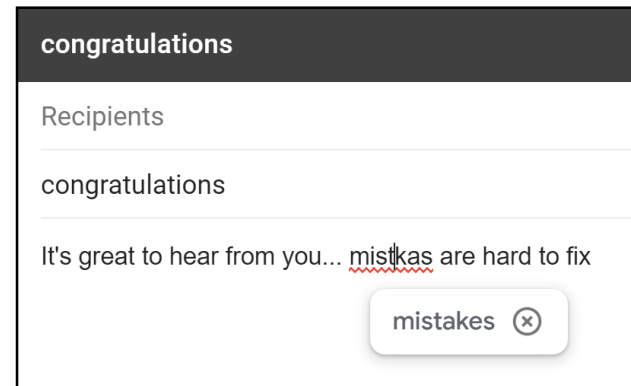
Navigation Choices



Texting/Search Autocompletion



Spelling correction



Supertool: Bloomberg Terminal



Active Appliances

Thermostat, Coffee maker, Rice cooker, Roomba



Google Nest



Panasonic Rice Cooker



Cuisinart Coffee Maker



iRobot Roomba

Dishwasher, Clothes Washer/Dryer



Miele Dishwasher

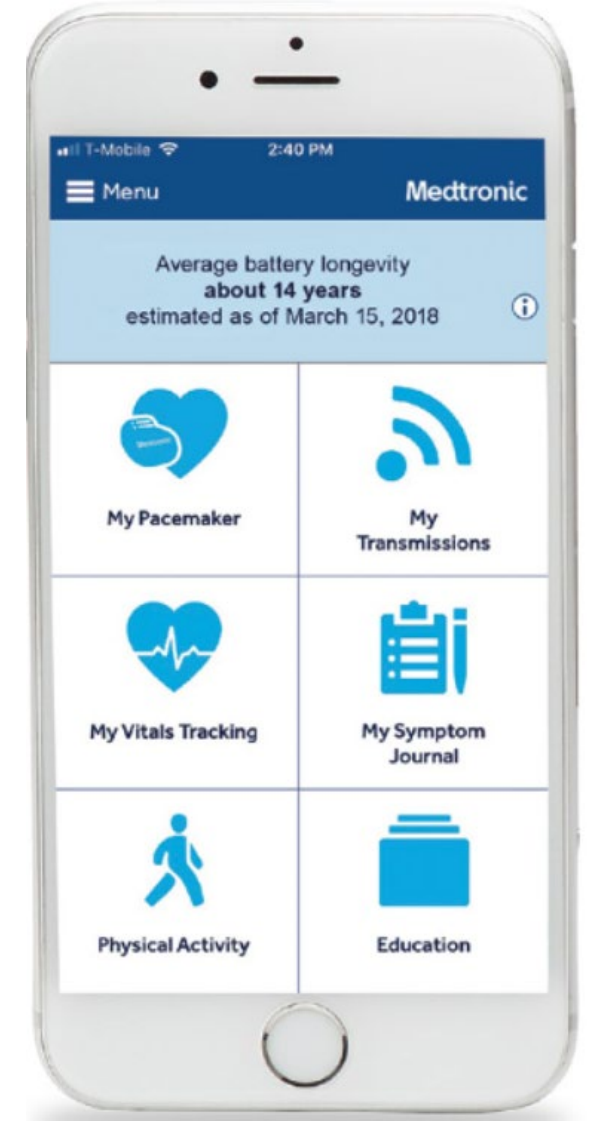
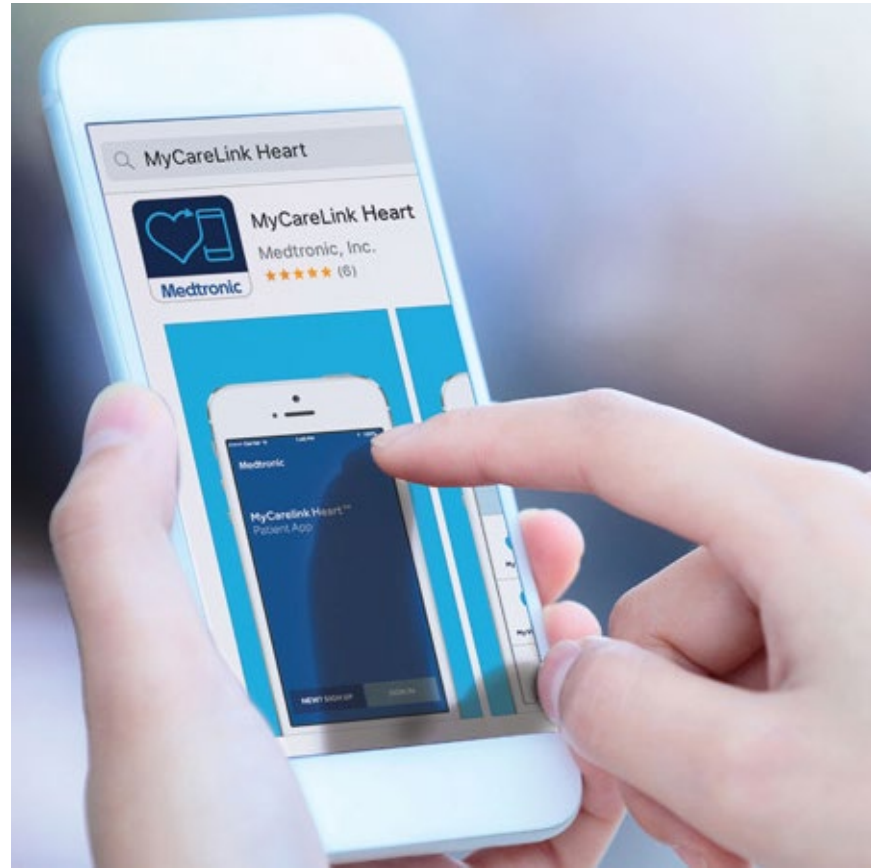
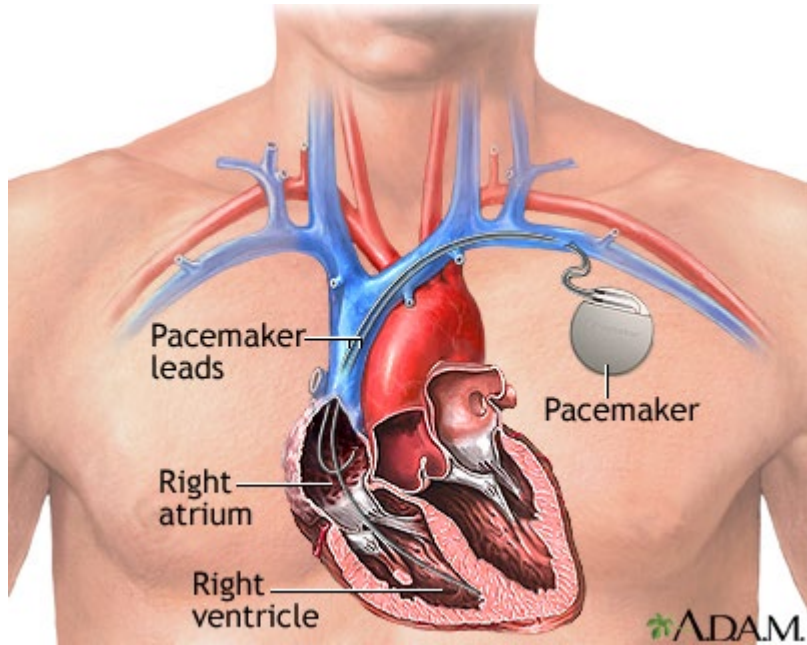


General Electric Washer



General Electric Dryer

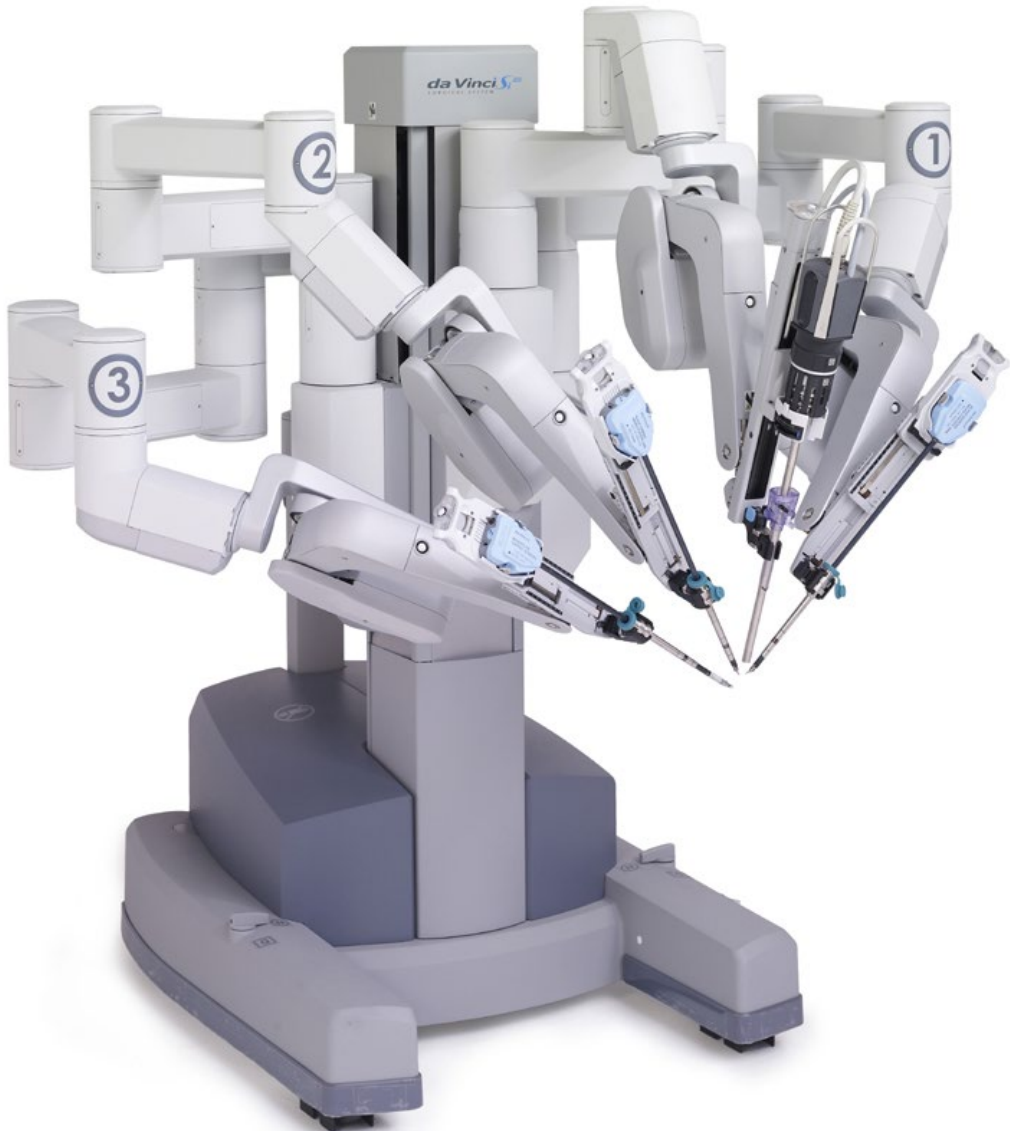
Active Appliance: Implanted Cardiac Pacemakers



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



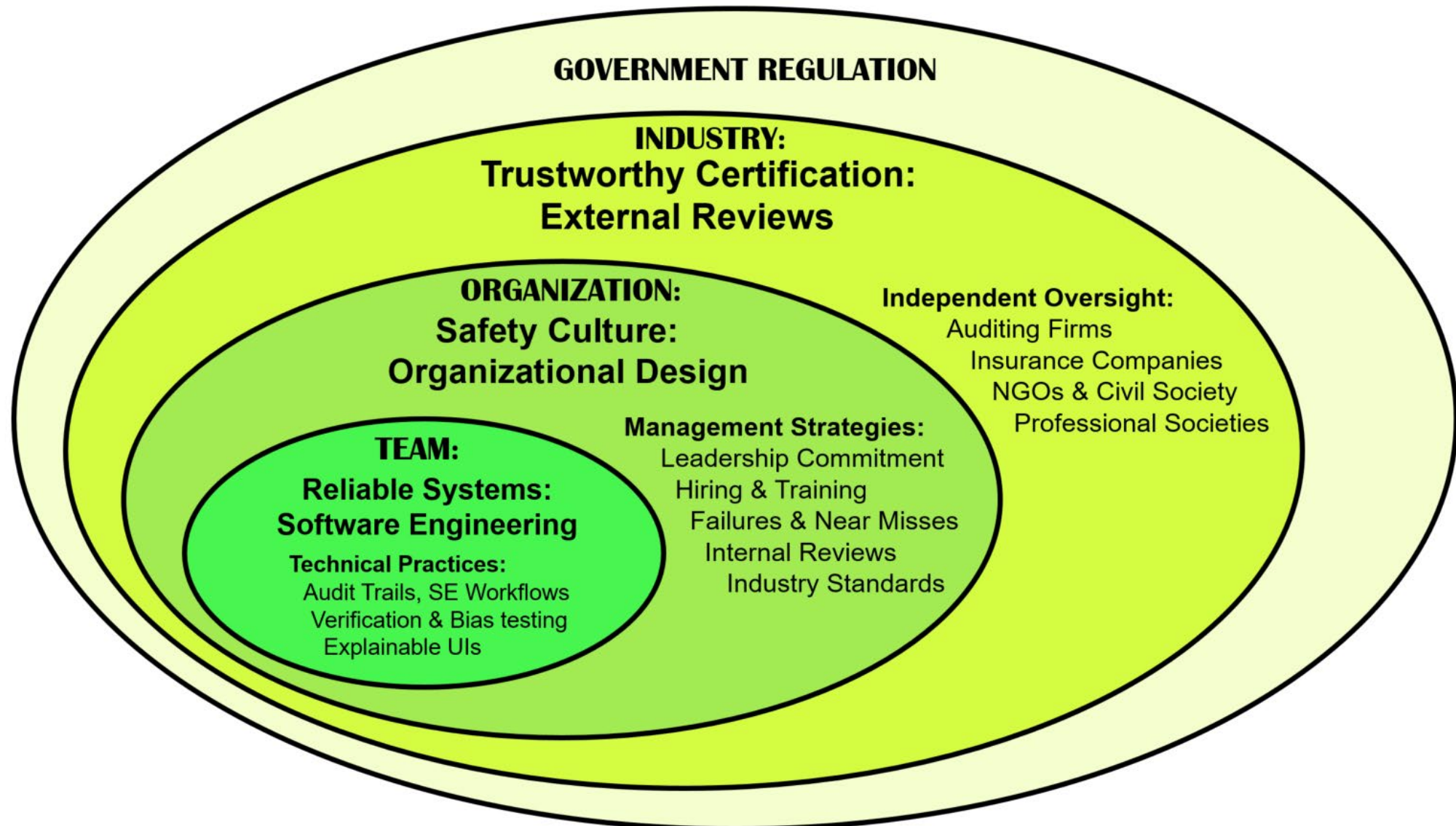
Control Center: Counter Terrorism



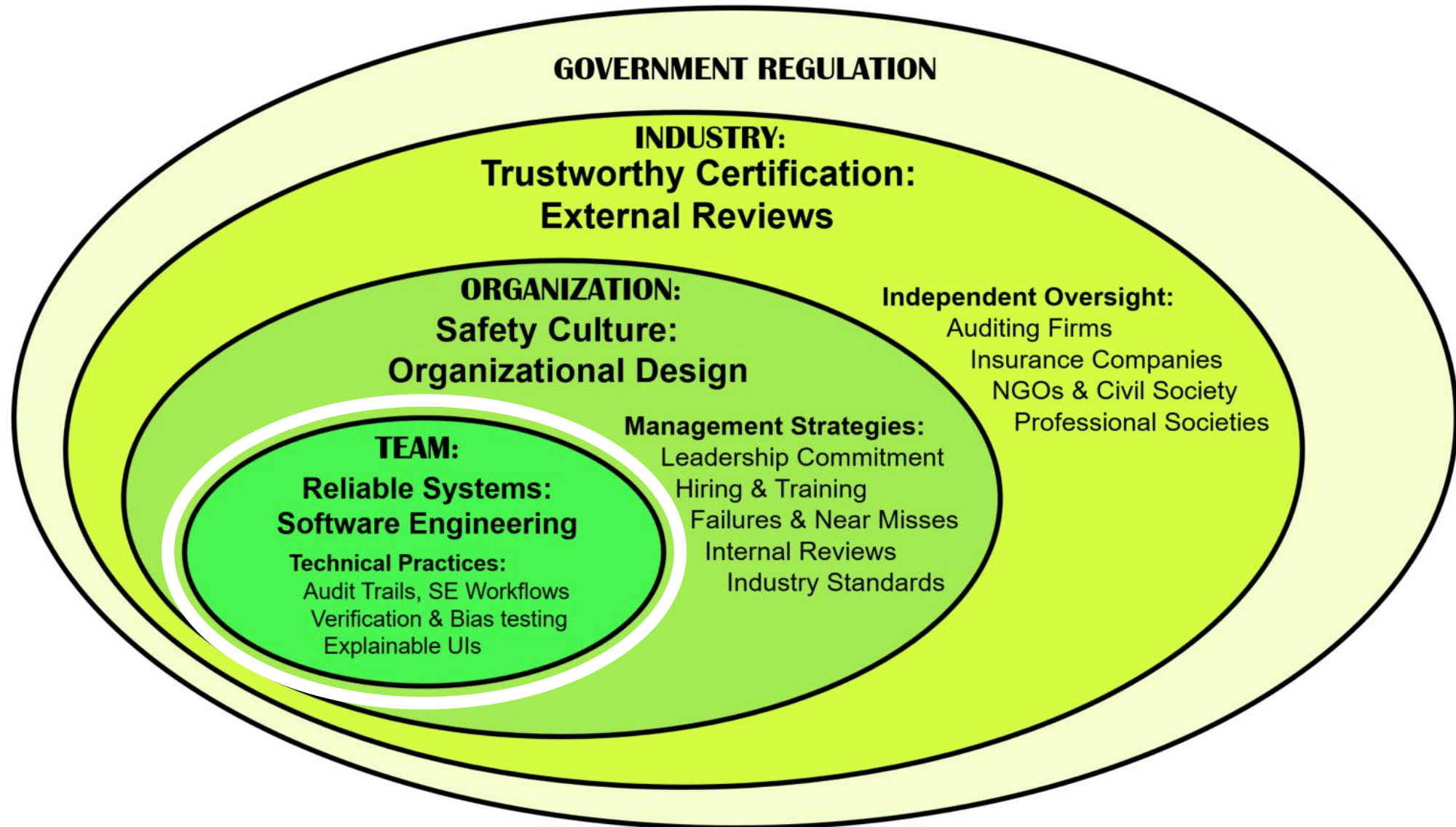
Governance Structures



Governance Structures for Human-Centered AI



Governance Structures for Human-Centered AI



TEAM

Reliable systems based on software engineering practices

- 1) Audit trails and analysis tools
- 2) Software engineering workflows
- 3) Verification & validation testing
- 4) Bias testing to improve fairness
- 5) Explainable user interfaces

TEAM

Reliable systems based on software engineering practices

- 1) Audit trails and analysis tools
- 2) Software engineering workflows
- 3) Verification & validation testing
- 4) Bias testing to improve fairness
- 5) Explainable user interfaces

Reliable Systems

Software engineering practices for a TEAM

5) Explainable user interfaces

- Retrospective explanations (local & global)

New Goal: **Prevent** confusion and surprise

- Prospective user interfaces
- Interactive, visual, exploratory

Mortgage Loan Explanations

Post-hoc Report

Enter amounts to request mortgage:

Mortgage amount requested

Household monthly income

Liquid assets

Mortgage Loan Explanations

Post-hoc Report

Enter amounts to request mortgage:

Mortgage amount requested	<input type="text" value="375000"/>
Household monthly income	<input type="text" value="7000"/>
Liquid assets	<input type="text" value="48000"/>

Enter amounts to request mortgage:

Mortgage amount requested	<input type="text" value="375000"/>
Household monthly income	<input type="text" value="7000"/>
Liquid assets	<input type="text" value="48000"/>

We're sorry, your mortgage loan was not approved. You might be approved if you reduce the Mortgage amount requested, increase your Household monthly income, or increase your Liquid assets.

Mortgage Loan Explanations

Post-hoc Report

Enter amounts to request mortgage:

Mortgage amount requested

Household monthly income

Liquid assets

Enter amounts to request mortgage:

Mortgage amount requested

Household monthly income

Liquid assets

We're sorry, your mortgage loan was not approved. You might be approved if you reduce the Mortgage amount requested, increase your Household monthly income, or increase your Liquid assets.

Prospective User Interface

Adjust sliders to report your situation:

Mortgage amount requested

Household monthly income

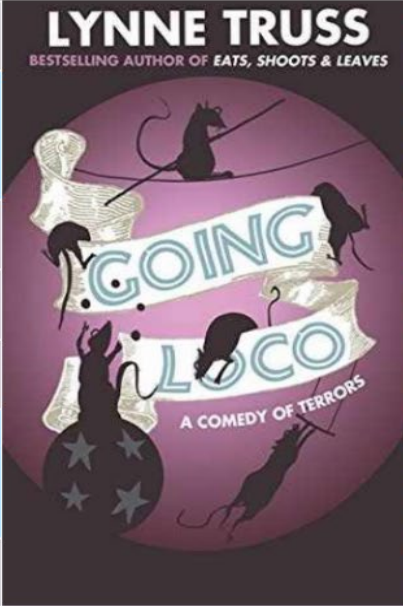
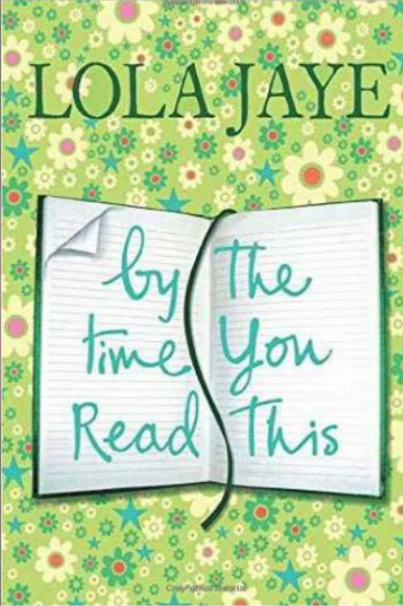
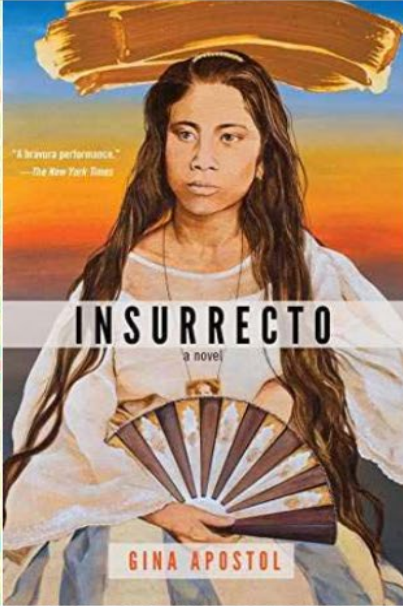
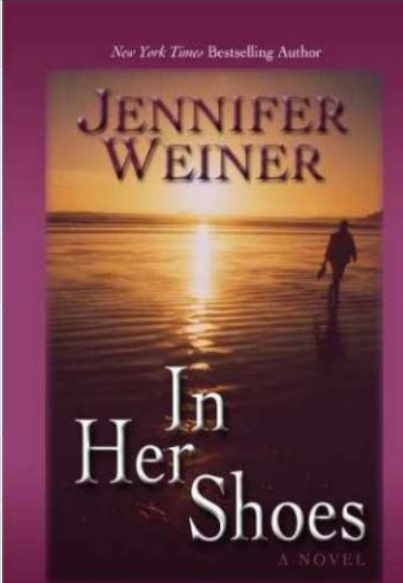
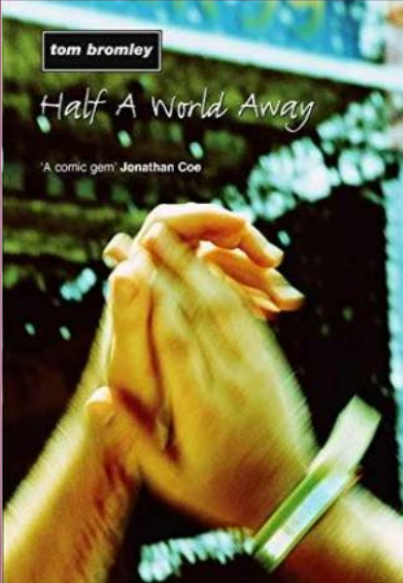

Liquid assets

Score needed for approval

Your score

Parameter	Value
Mortgage amount requested	375000
Household monthly income	7000
Liquid assets	48000

Recommenders: Whichbook.net

Happy <input type="checkbox"/> Sad			
Funny <input checked="" type="checkbox"/> Serious			
Safe <input type="checkbox"/> Disturbing			
Expected <input type="checkbox"/> Unpredictable			
Larger than life <input type="checkbox"/> Down to earth			
Beautiful <input checked="" type="checkbox"/> Disgusting			
Gentle <input type="checkbox"/> Violent			
Easy <input type="checkbox"/> Demanding			
No sexual content <input checked="" type="checkbox"/> Explicit sexual content			
Conventional <input type="checkbox"/> Unusual			
Optimistic <input checked="" type="checkbox"/> Bleak			
Short <input type="checkbox"/> Long			
Select up to 4 sliders			

Recommender Control Panels

Modify Attributes

acousticness: 40

instrumentalness: 60

danceability: 80

valence: 60

energy: 40

Slider technique

Recommended Songs

Calculate Recommendations

Task: Make a playlist of songs to listen to during your personal maintenance.

Click or to keep or dismiss a song.
More info:

- Oblivion
Grimes
- My December
Linkin Park
- I've got that tune
Chinese Man
- Got the Life
Korn
- Good Riddance (Time of Y...
Green Day
- Burn It To The Ground
Nickelback

To get more songs, modify the attribute(s) and click "Calculate Recommendations" again.

Create Your Better Life Index

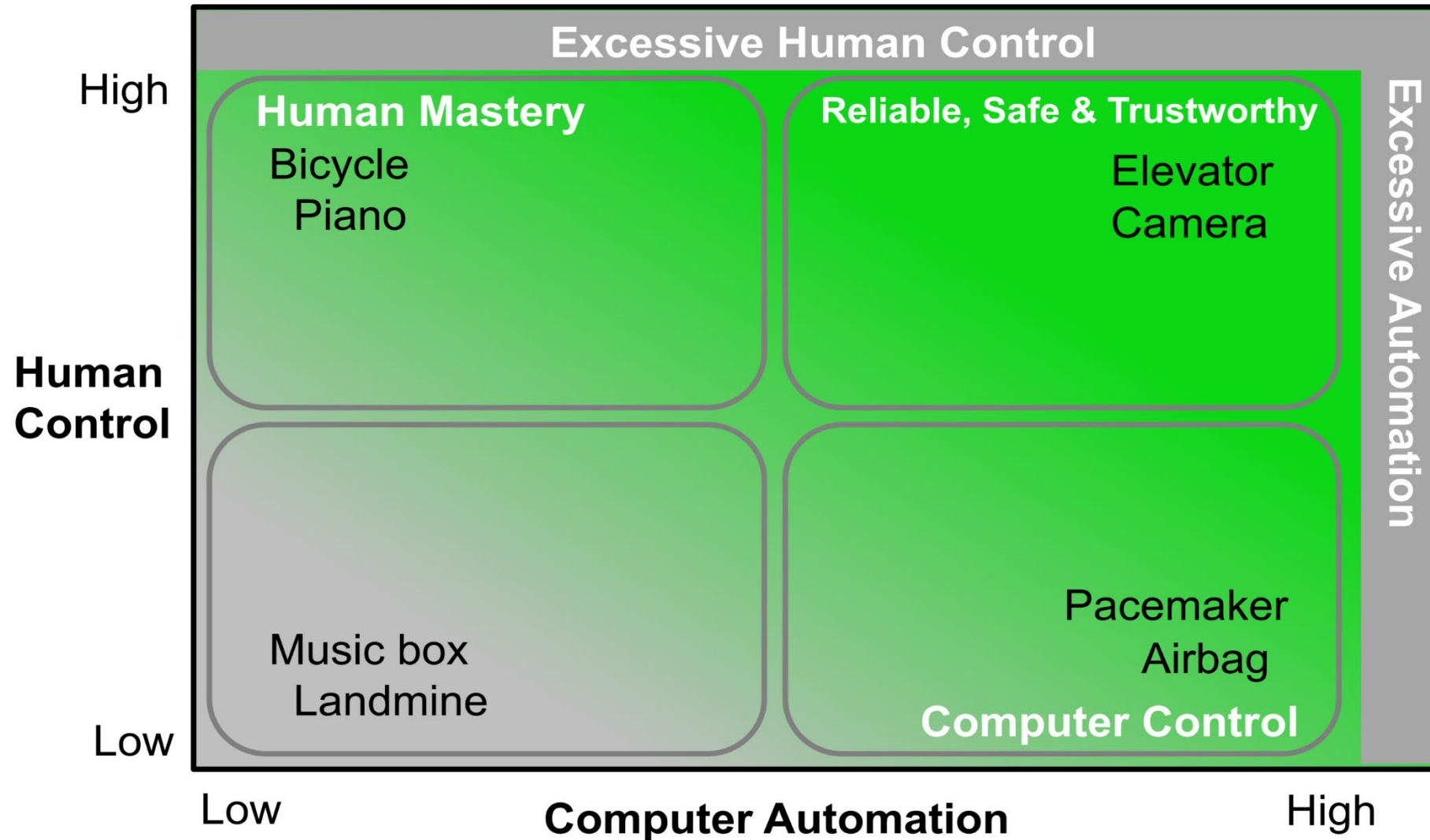
Rate the topics according to their importance to you:

	-	+
Housing		
Income		
Jobs		
Community		
Education		
Environment		
Civic Engagement		
Health		
Life Satisfaction		
Safety		
Work-Life Balance		

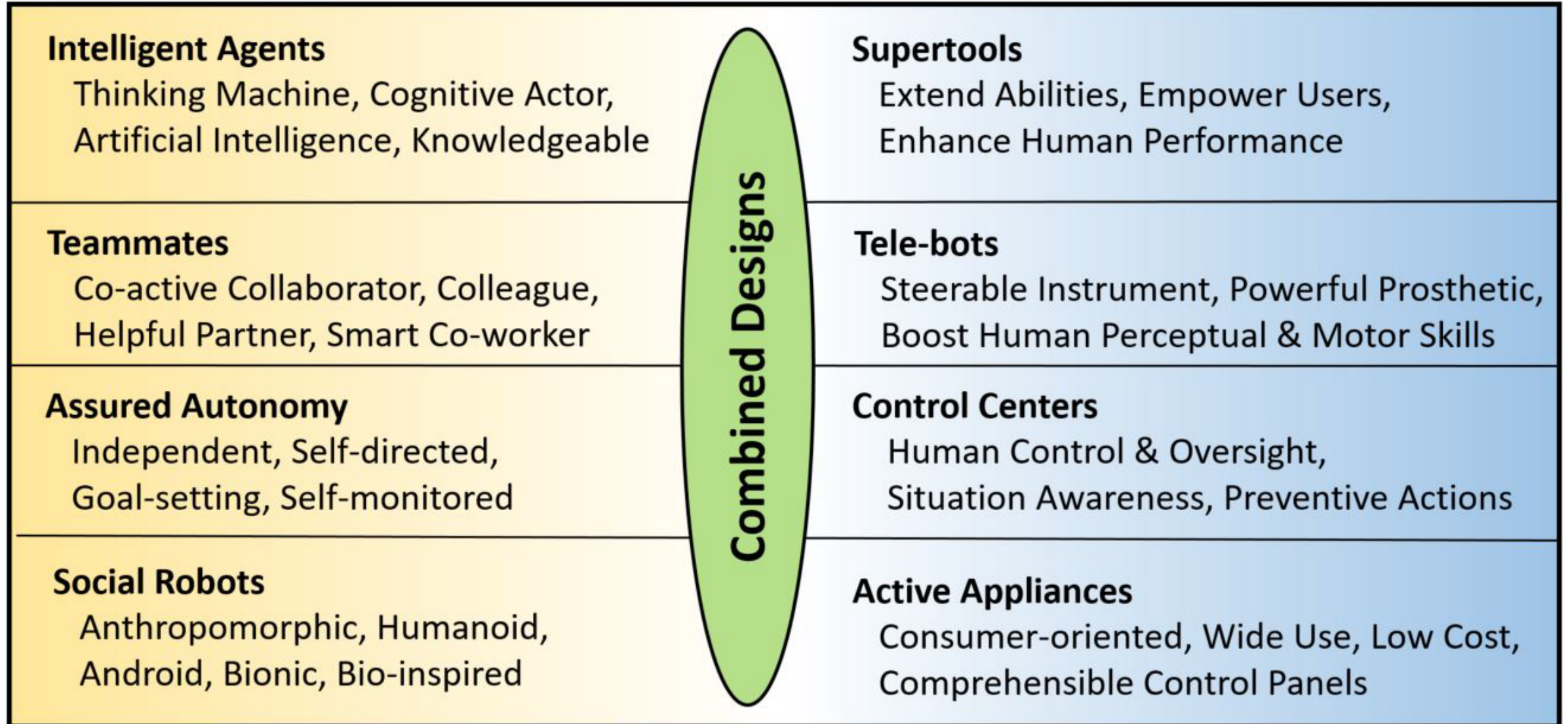


Summary

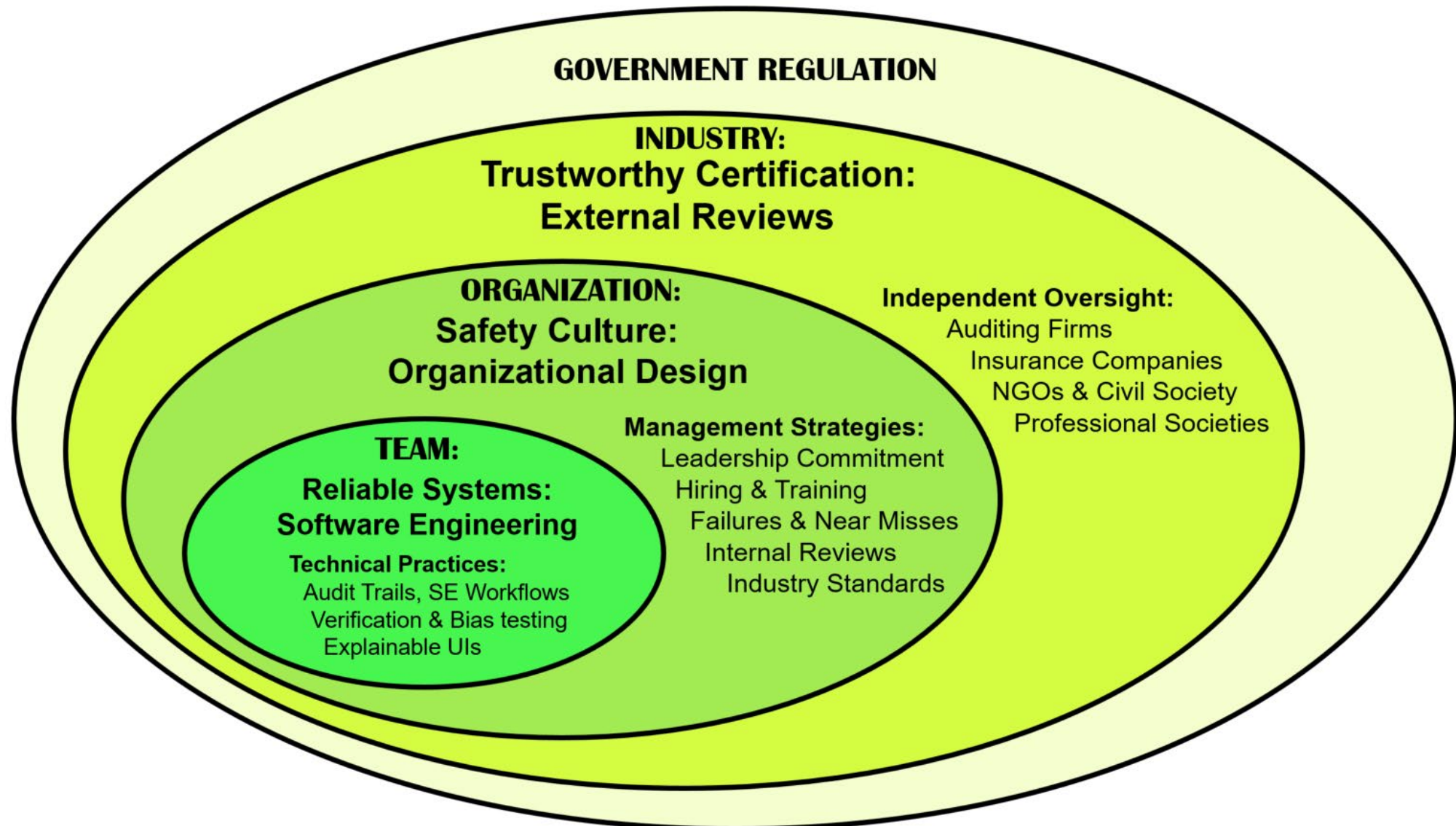
HCAI Framework



Design Metaphors



Governance Structures for Human-Centered AI



Human-Centered AI

Oxford University Press

- January 2022

The book is well-structured and a delight to read. The coverage is comprehensive. But it will be controversial.

— Alan Mackworth, University of British Columbia

From design metaphors to the much needed governance structures, this new book by Ben Shneiderman is a tour de force into the increasingly important topic of human-centred AI. A must read.

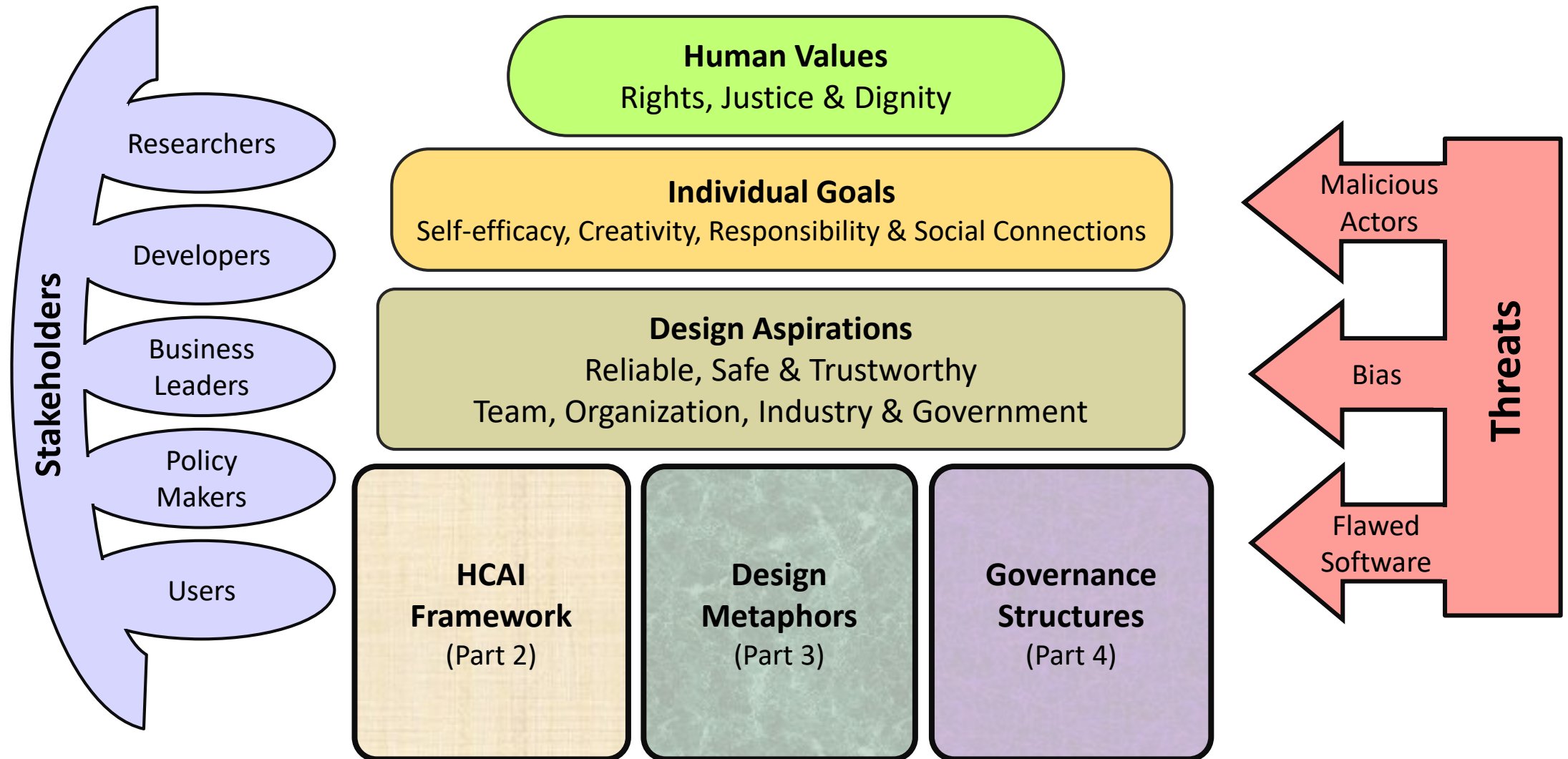
— Virginia Dignum, Umeå University"

Intellectually re-positioning the practice of AI is the most important social movement of our age. Human-Centered AI is a moral imperative.

— Sean McGregor, Founder and Project Lead, the AI Incident Database
(Partnership on AI)



Human-Centered AI



The Future is Human-Centered

Google Group

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

Twitter Account

@HumanCenteredAI

Wikipedia Page

https://en.wikipedia.org/wiki/Human-Centered_Artificial_Intelligence

Website

<https://hcai.site>

The Future is Human-Centered



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/>