

Arizona State University, DC, March 13, 2020

Human-Centered Artificial Intelligence: 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

Web links

The **University of Maryland, College Park** (often referred to as the **University of Maryland, Maryland, UM, UMD, UMCP**, or **College Park**) is a public research university^[10] located in the city of **College Park** in **Prince George's County, Maryland**, approximately 4 miles (6.4 km) from the northeast border of Washington, D.C. Founded in 1856, the university is the **flagship** institution of the **University System of Maryland**. With a fall 2010 enrollment of more than 37,000 students, over 100 undergraduate majors, and 120 graduate programs,

Tiny touchscreen keyboards

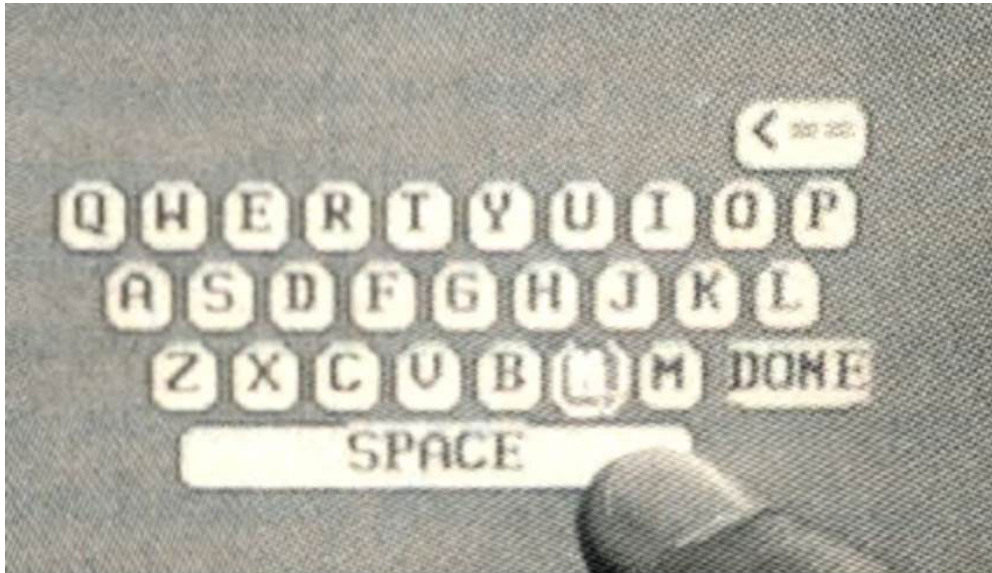
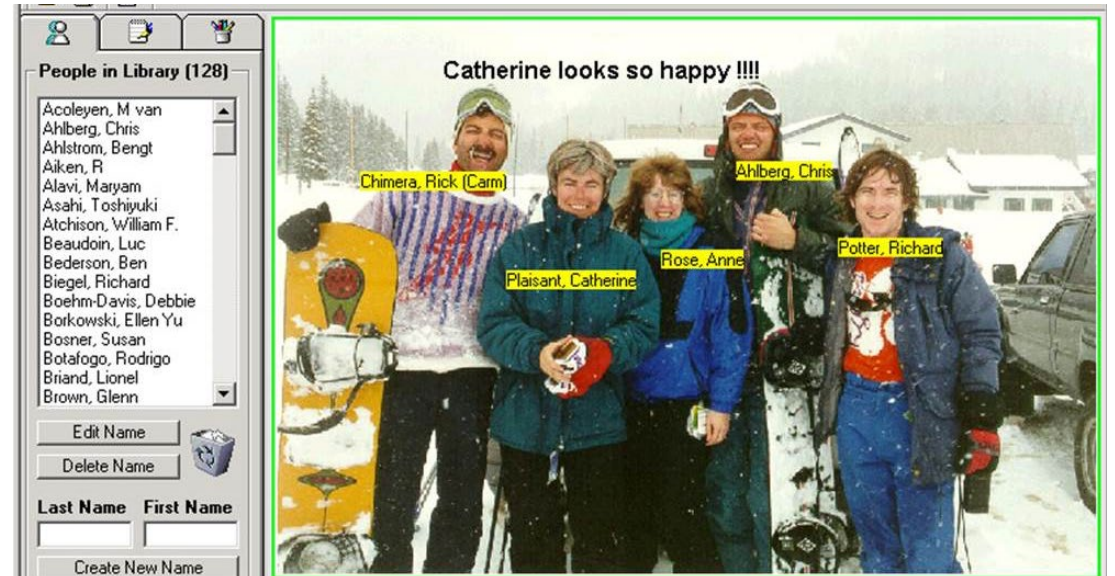
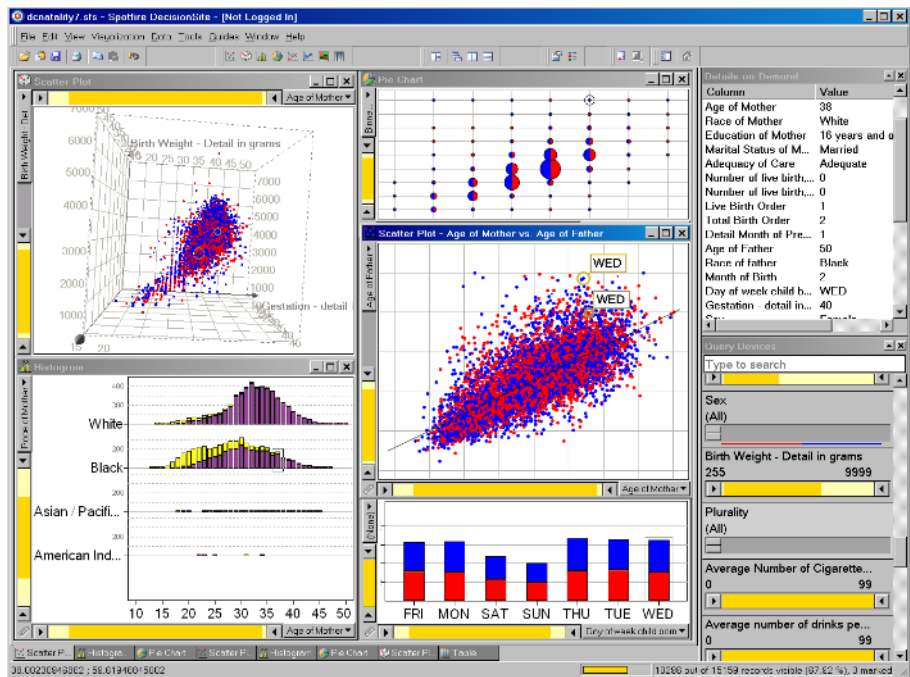


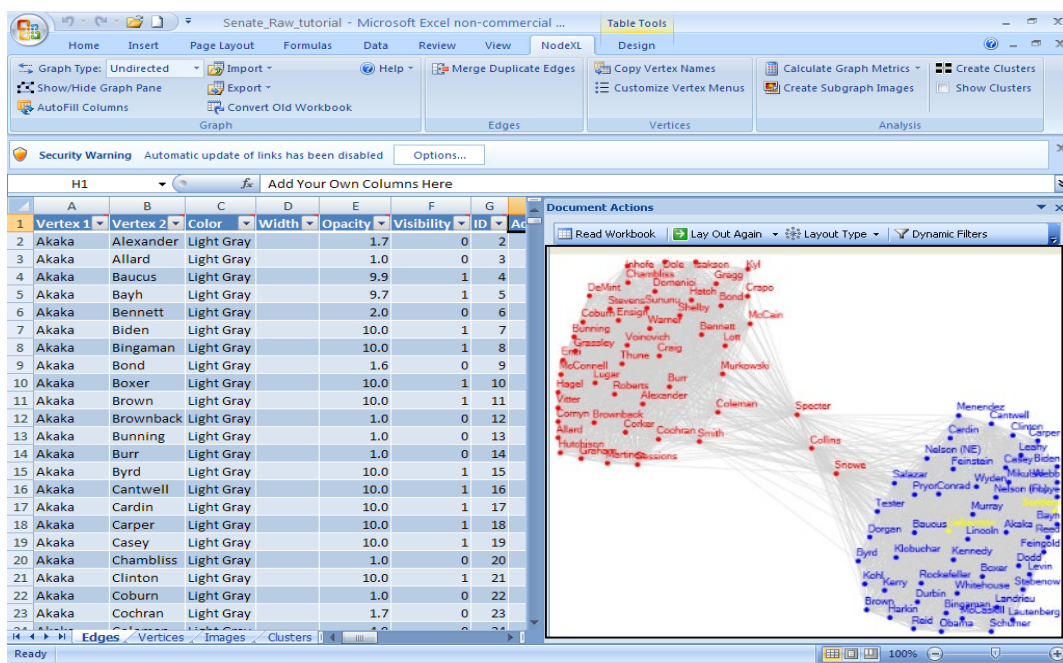
Photo tagging



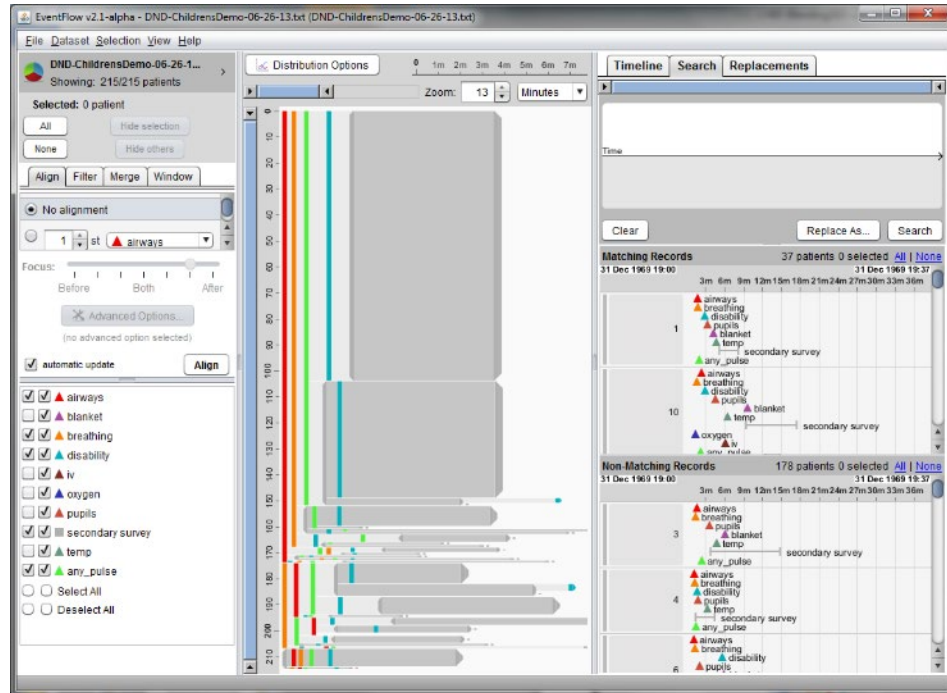


Spotfire

Treemaps
FinViz



NodeXL



EventFlow



Rationalism vs Empiricism

Rationalism: logical perfectability of human endeavors

Aristotle, Descartes, Spinoza & Kant

Empiricism: need to continuously refine our beliefs
based on our experiences

Leonardo, Galileo, Locke & Hume

Rationalism vs Empiricism

Rationalism: logical perfectability of human endeavors

Aristotle, Descartes, Spinoza & Kant ... Ronald Fisher

Empiricism: need to continuously refine our beliefs
based on our experiences

Leonardo, Galileo, Locke & Hume ... John Tukey

Rationalism vs Empiricism

Rationalism: logical perfectability of human endeavors

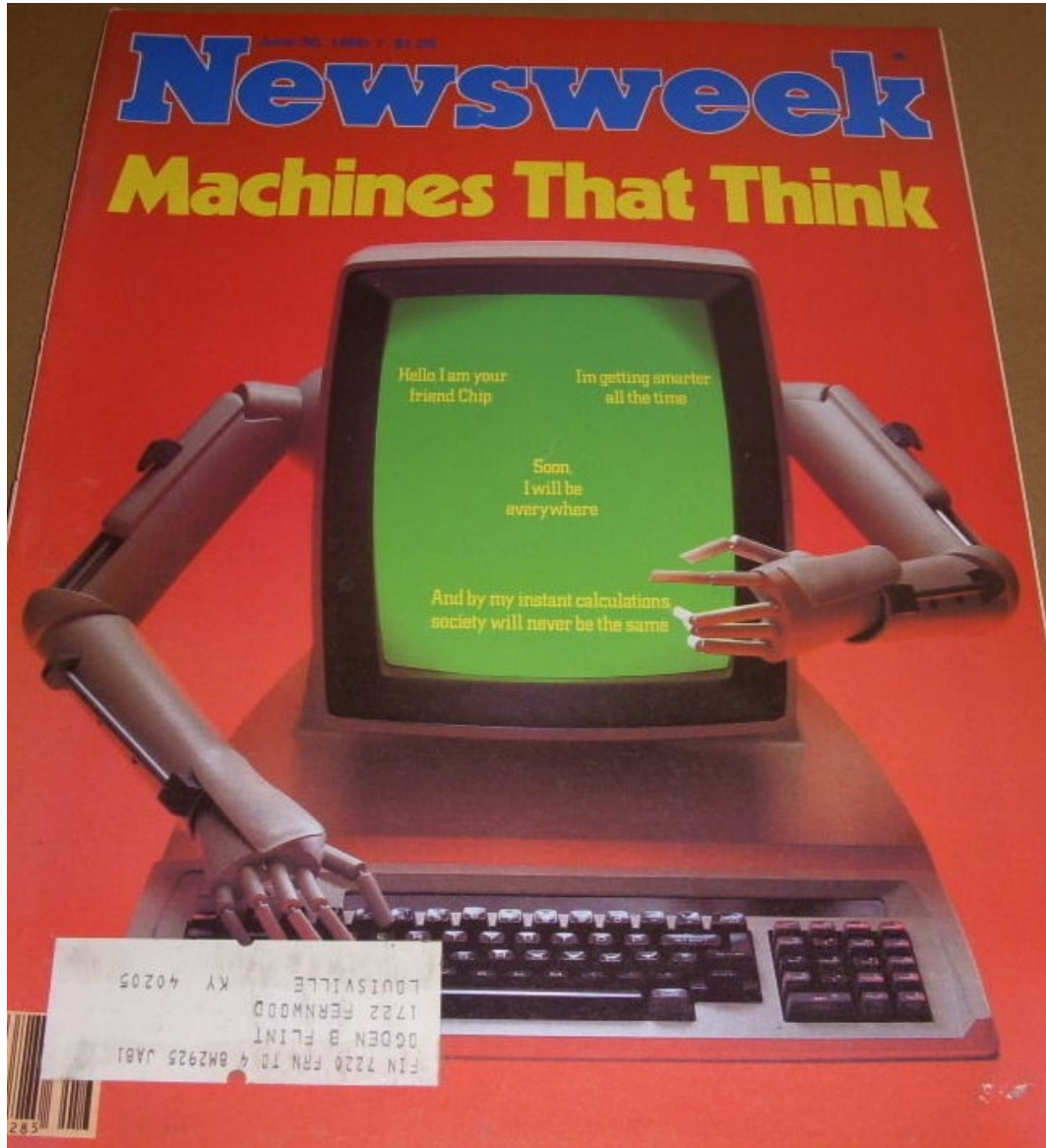
Aristotle, Descartes, Spinoza & Kant ... AI

Empiricism: need to continuously refine our beliefs
based on our experiences

Leonardo, Galileo, Locke & Hume ... HCI

Machines That Think

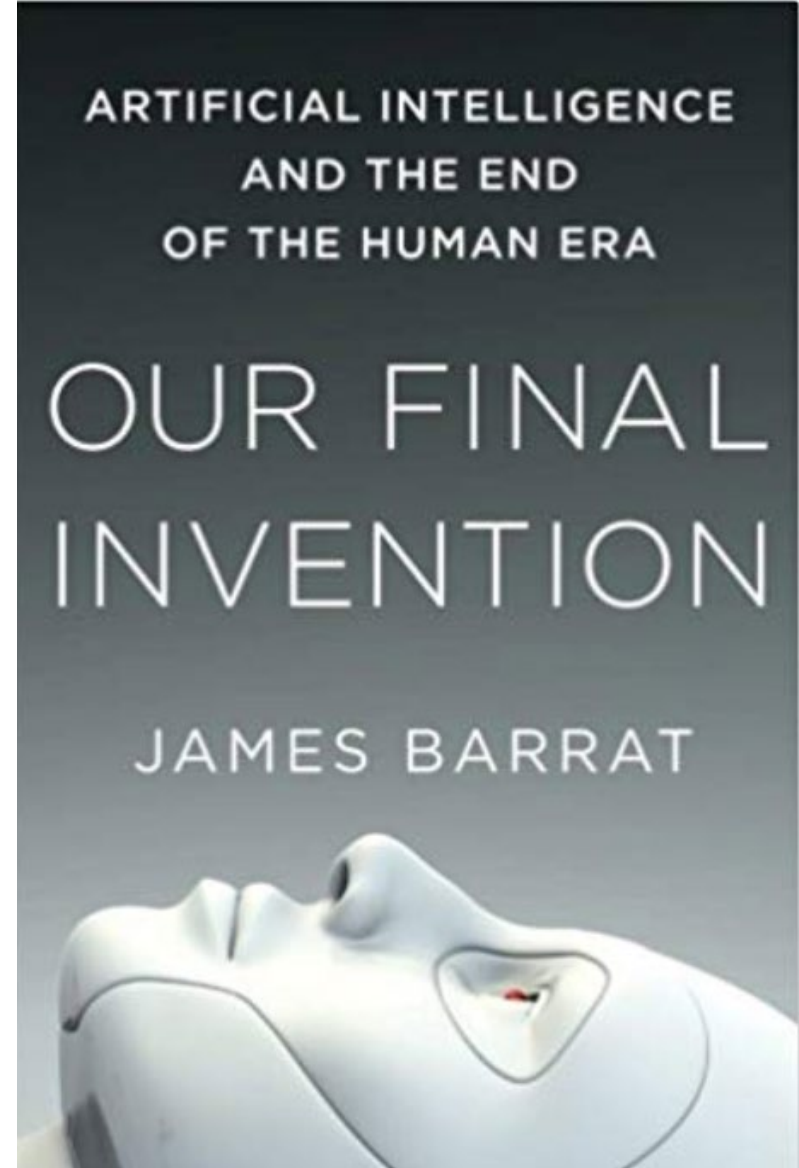
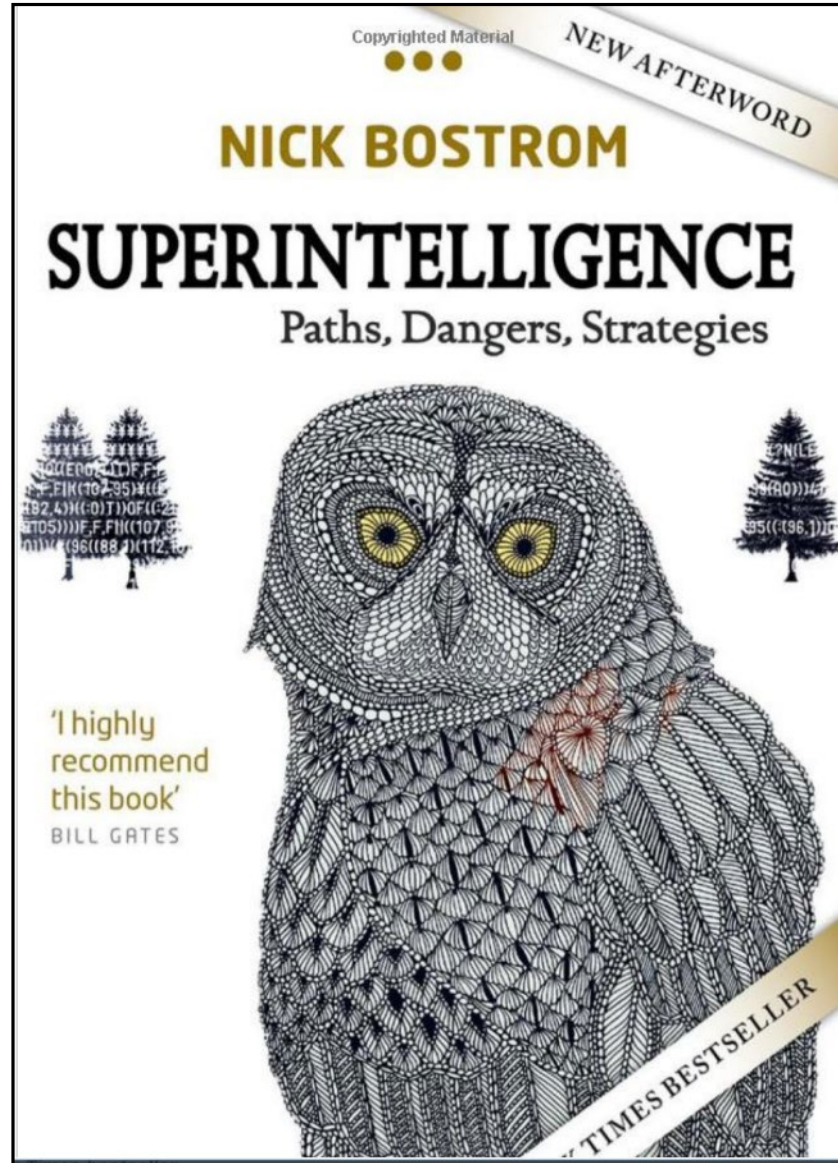
June 30, 1980



Can Machines Think?

March 25, 1996





Copyrighted Material

WEAPONS OF MATH DESTRUCTION



HOW BIG DATA INCREASES INEQUALITY
AND THREATENS DEMOCRACY

CATHY O'NEIL

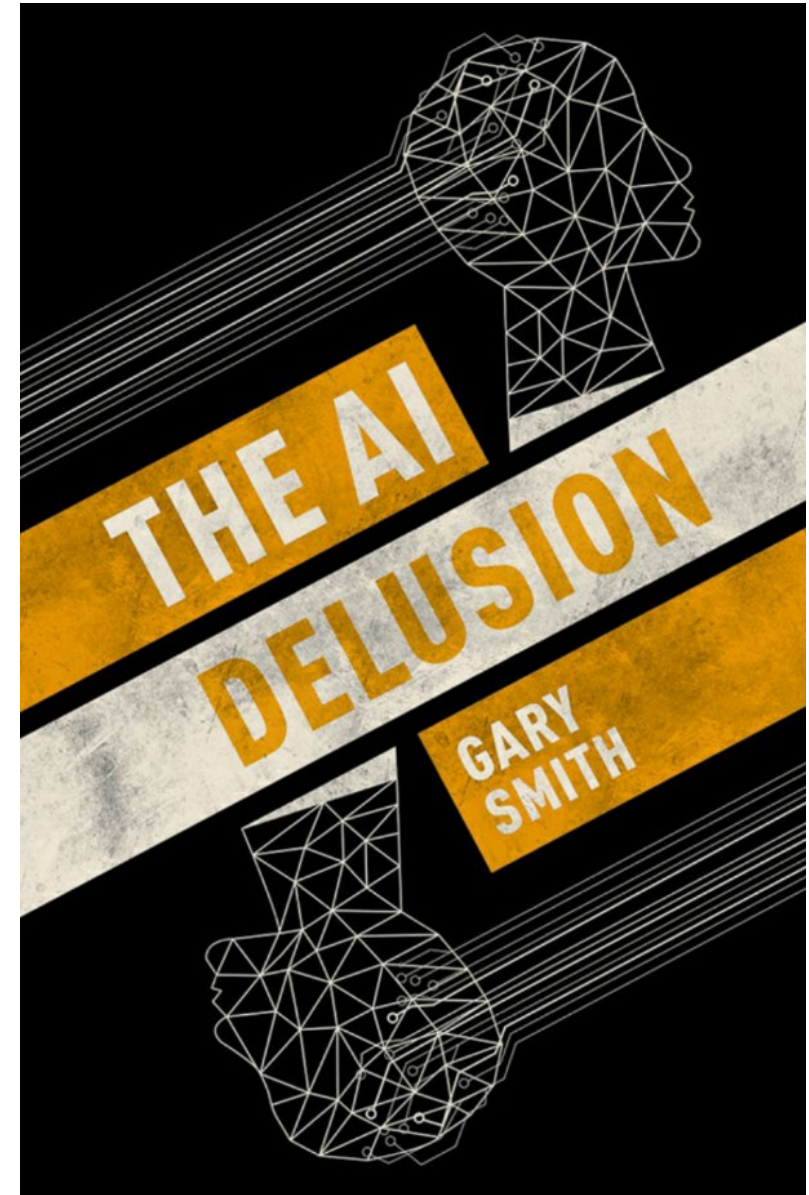
Copyrighted Material



REBOOTING AI

Building Artificial
Intelligence We Can Trust

GARY MARCUS and ERNEST DAVIS



Principled Artificial Intelligence: Governance

**BERKMAN
KLEIN CENTER**
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

Summarizes
36 leading reports
on AI Governance

- Privacy
- Accountability
- Safety & security
- Transparency & explainability
- Fairness & non-discrimination
- Human control of technology
- Professional responsibility
- Promotion of human values
- **International human rights**

HCI Pride: Serving 6 Billion Users

Mobile, desktop, web, cloud

Diverse users:

include everyone

Diverse applications:

real problems, users, data

Diverse interfaces:

theories, principles, guidelines

HCI Pride: Serving 6 Billion Users

Mobile, desktop, web, cloud

Diverse users:	include everyone
Diverse applications:	real problems, users, data
Diverse interfaces:	theories, principles, guidelines

Apple Design Guidelines

User Control

... people - not apps - are in control

Flexibility

... (give) users complete, fine-grained control over their work

AI design principles, which I want to change

- **Humanoid robots**
- **Excessive automation**
- Insufficient feedback
- Hidden features
- Restrictions on creativity
- Limits on collaboration
- Missing audit trails

Human-Centered AI

Amplify, Augment, Enhance & Empower People

Human-Centered AI

Amplify, Augment, Enhance & Empower People

→ 1000-fold improvements in capabilities

Information

Search

Email & Text

Photography

Navigation

Business Formation

Human-Centered AI

Amplify, Augment, Enhance & Empower People

→ 1000-fold improvements in capabilities

Information

Photography

Search

Navigation

Email & Text

Business Formation

→ RST: Reliable, Safe & Trustworthy

Human-Centered AI

Amplify, Augment, Enhance & Empower People

→ 1000-fold improvements in capabilities

Information

Photography

Search

Navigation

Email & Text

Business Formation

→ RST: Reliable, Safe & Trustworthy

→ More people, more creative, more often

Human-Centered AI

Amplify, Augment, Enhance & Empower People

→ 1000-fold improvements in capabilities

Information

Photography

Search

Navigation

Email & Text

Business Formation

→ RST: Reliable, Safe & Trustworthy

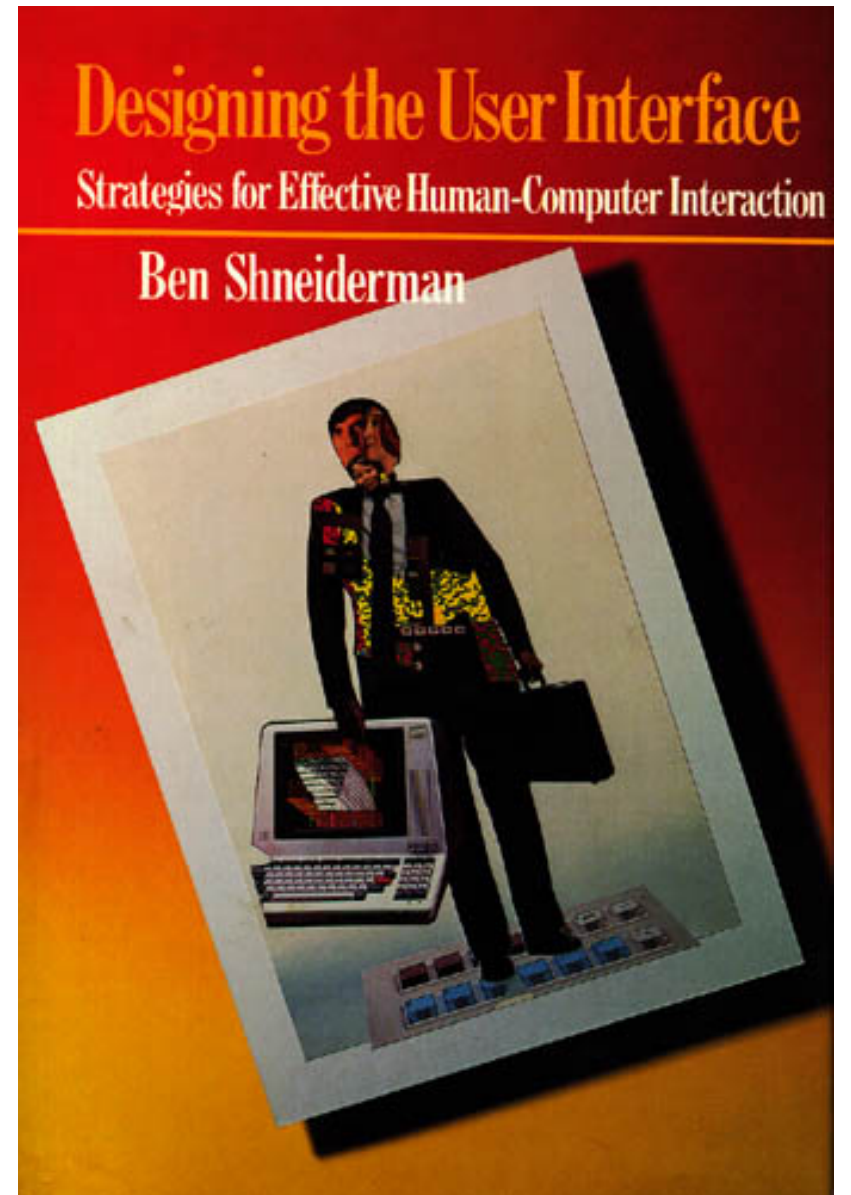
→ More people, more creative, more often

→ Self-efficacy & societal benefits



Designing the User Interface

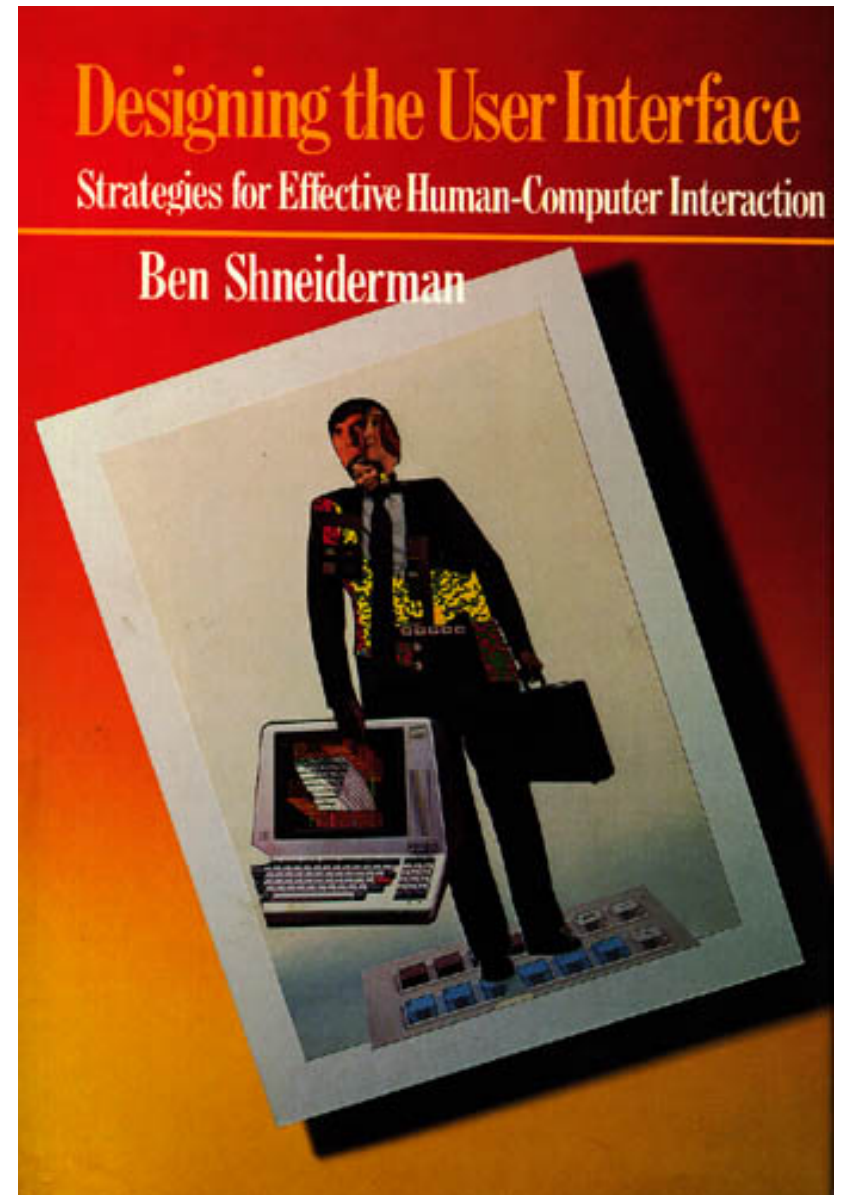
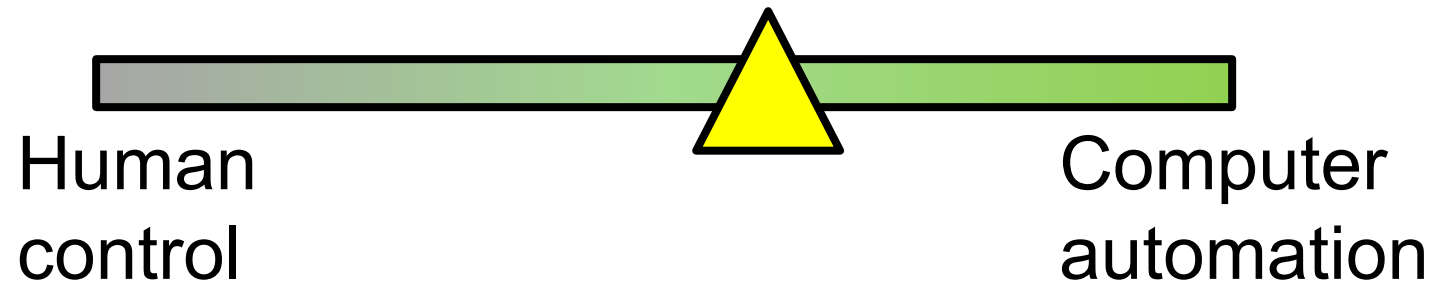
Balancing automation & human control



First Edition: 1986

Designing the User Interface

Balancing automation & human control



First Edition: 1986

Designing the User Interface

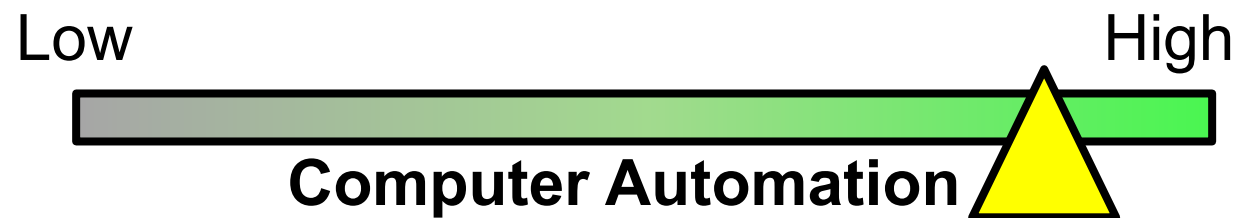
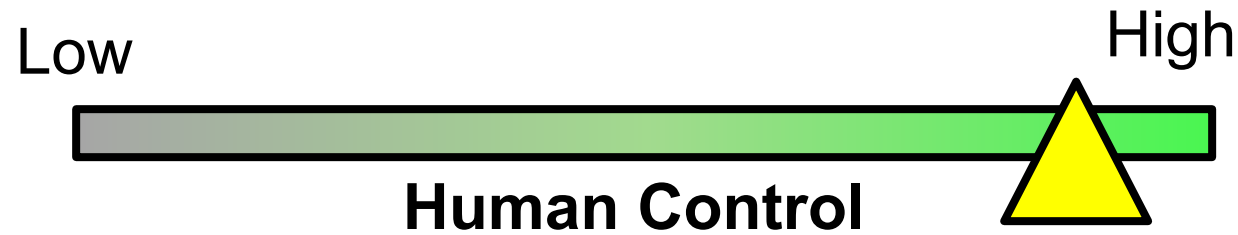
Ensuring human control
while increasing automation



Sixth Edition: 2016

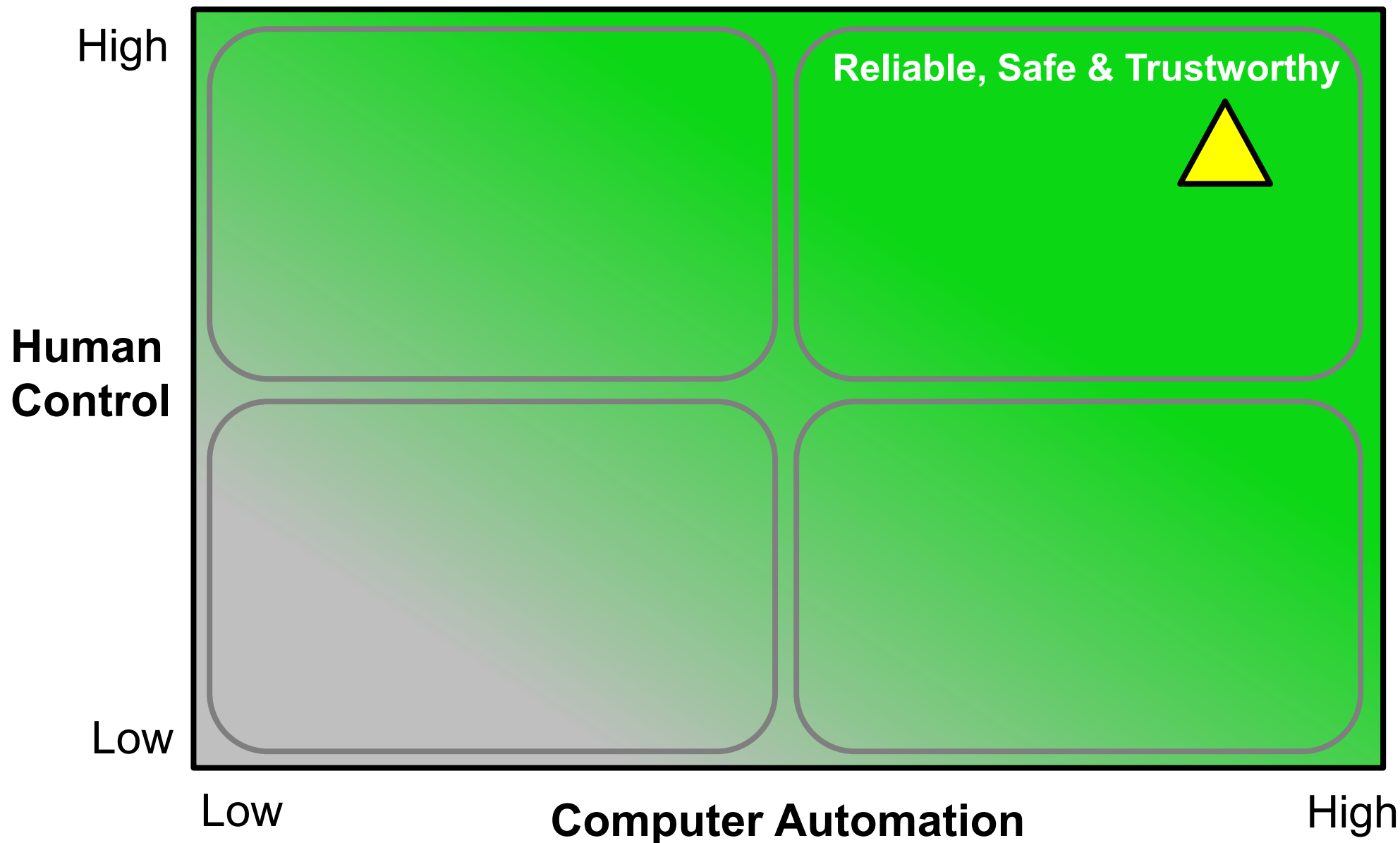
Designing the User Interface

Ensuring human control
while increasing automation

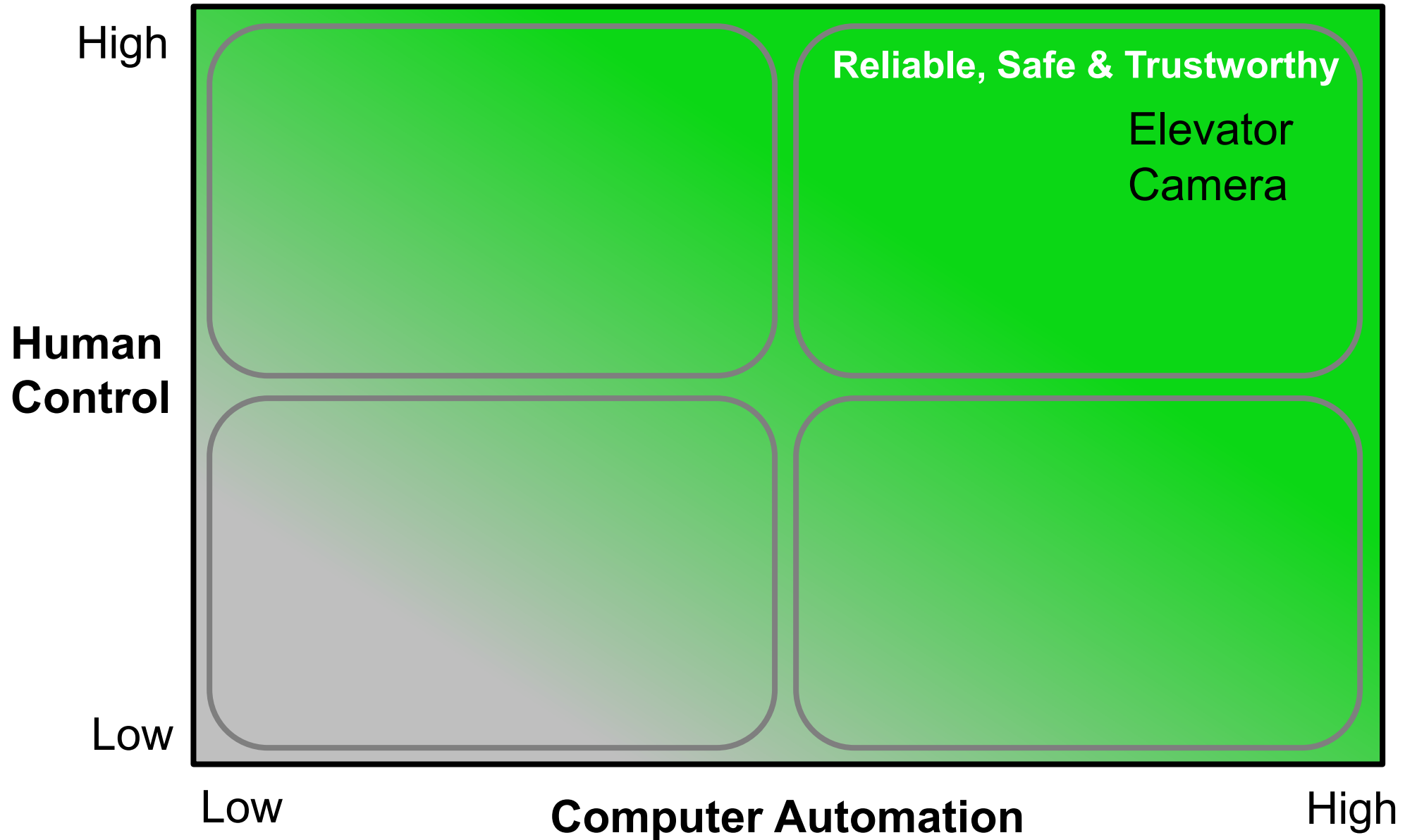


Sixth Edition: 2016

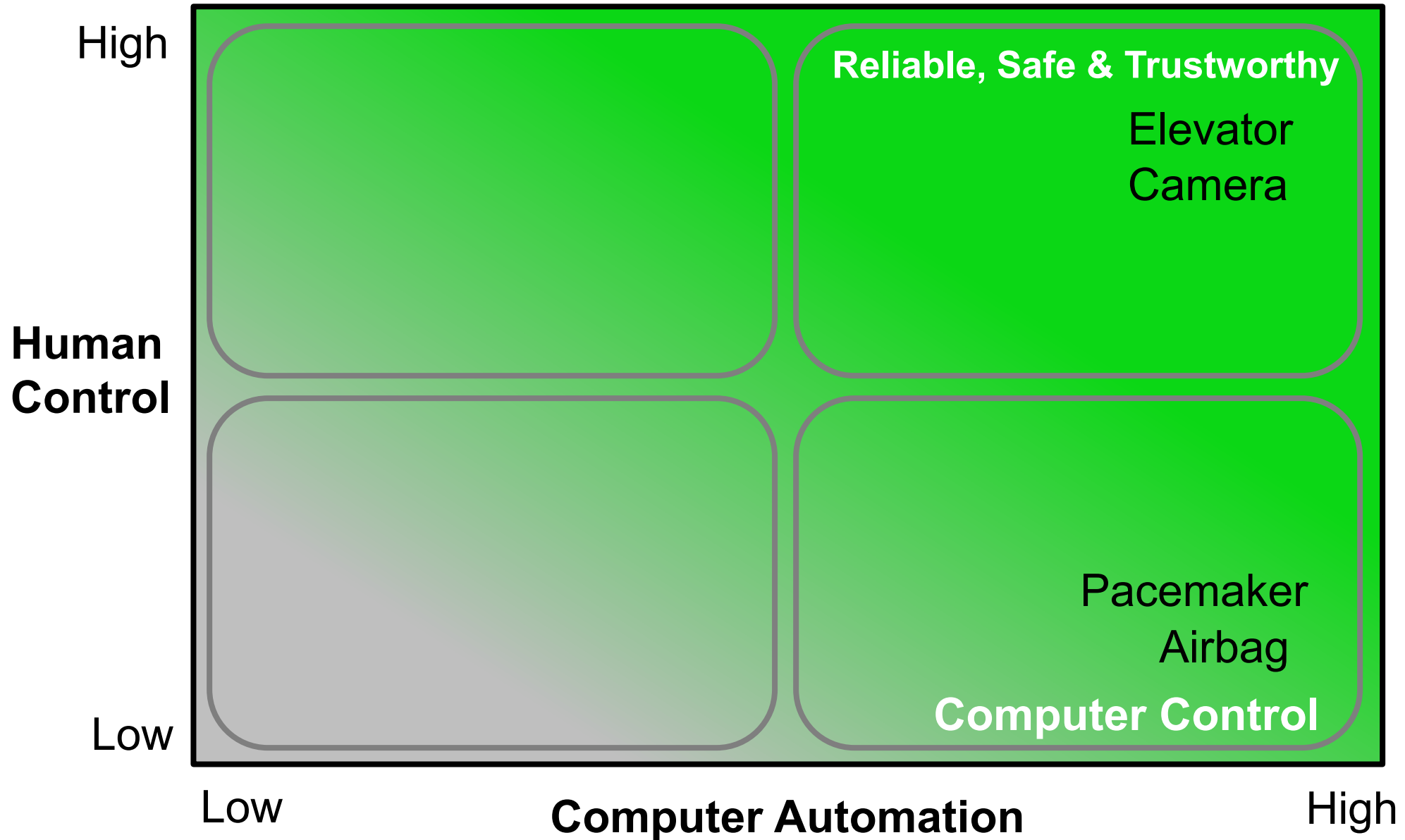
Human-Centered AI



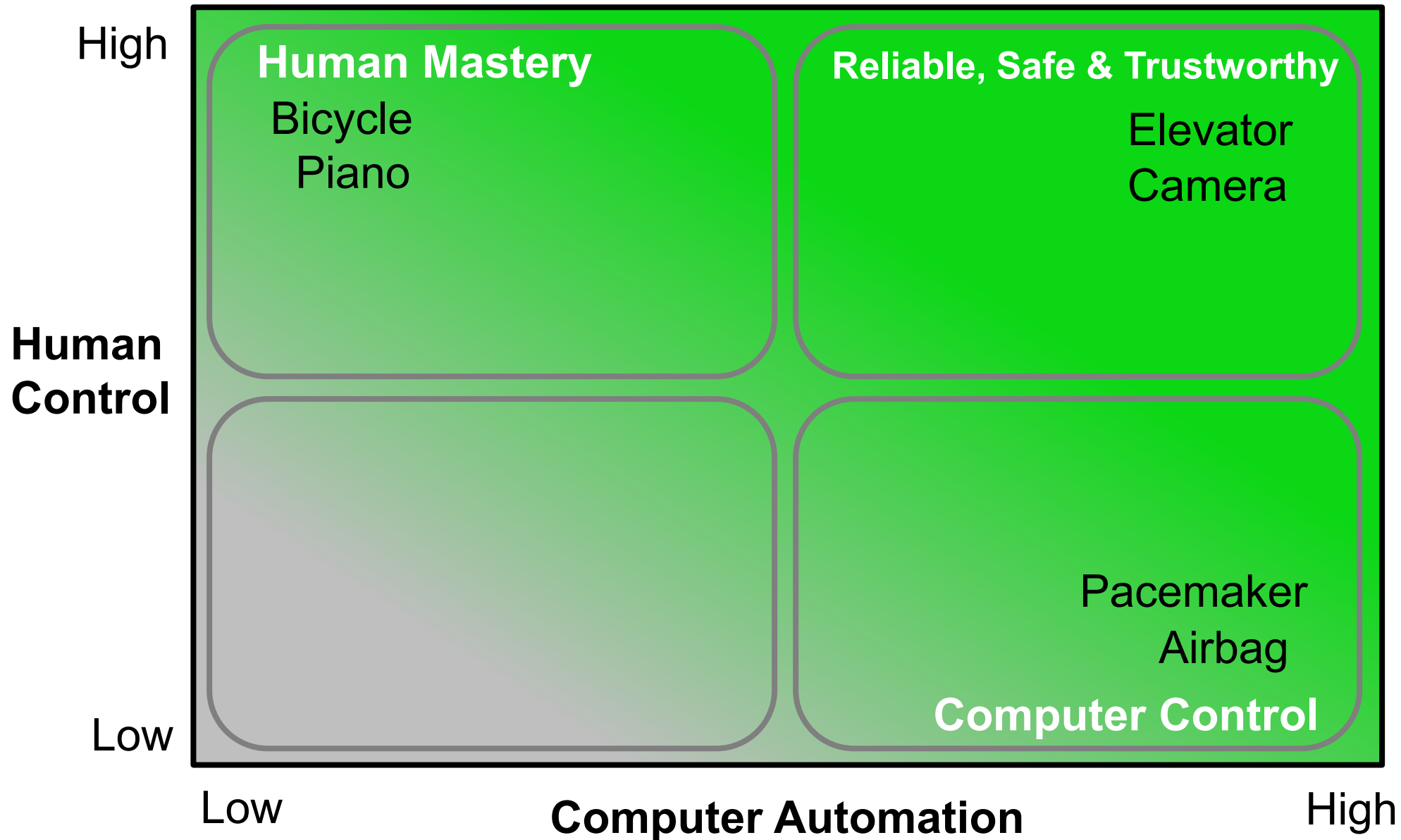
Human-Centered AI



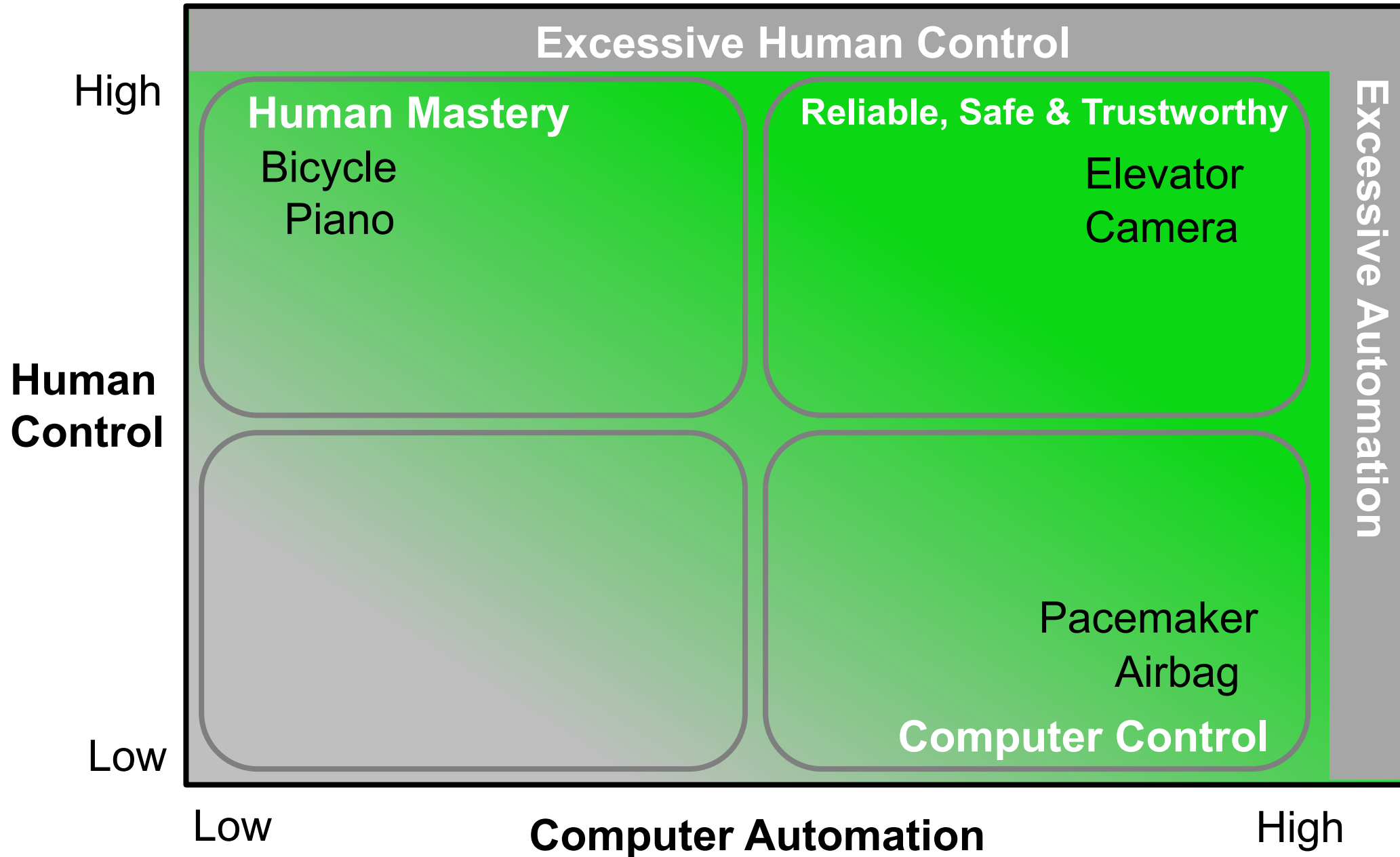
Human-Centered AI



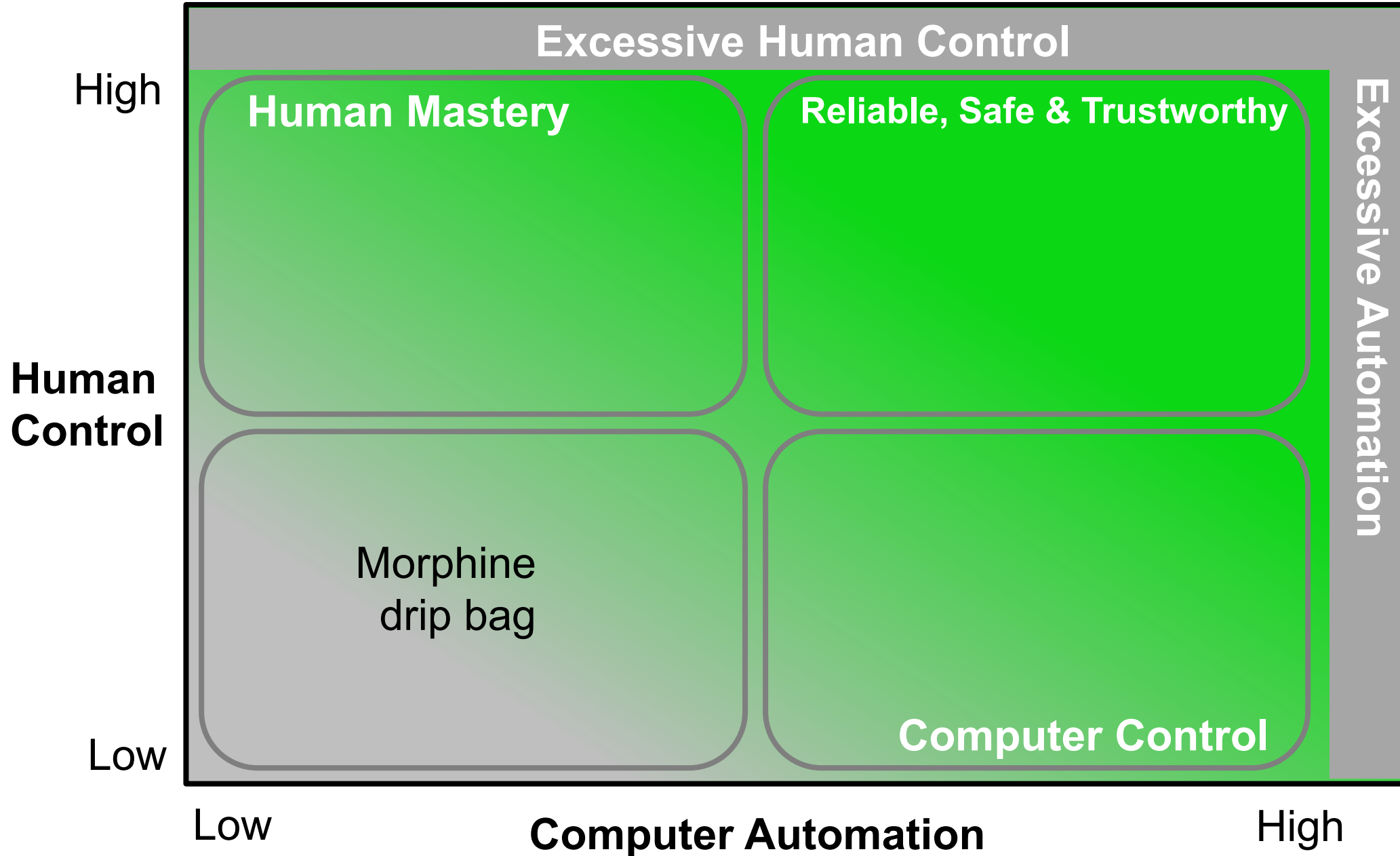
Human-Centered AI



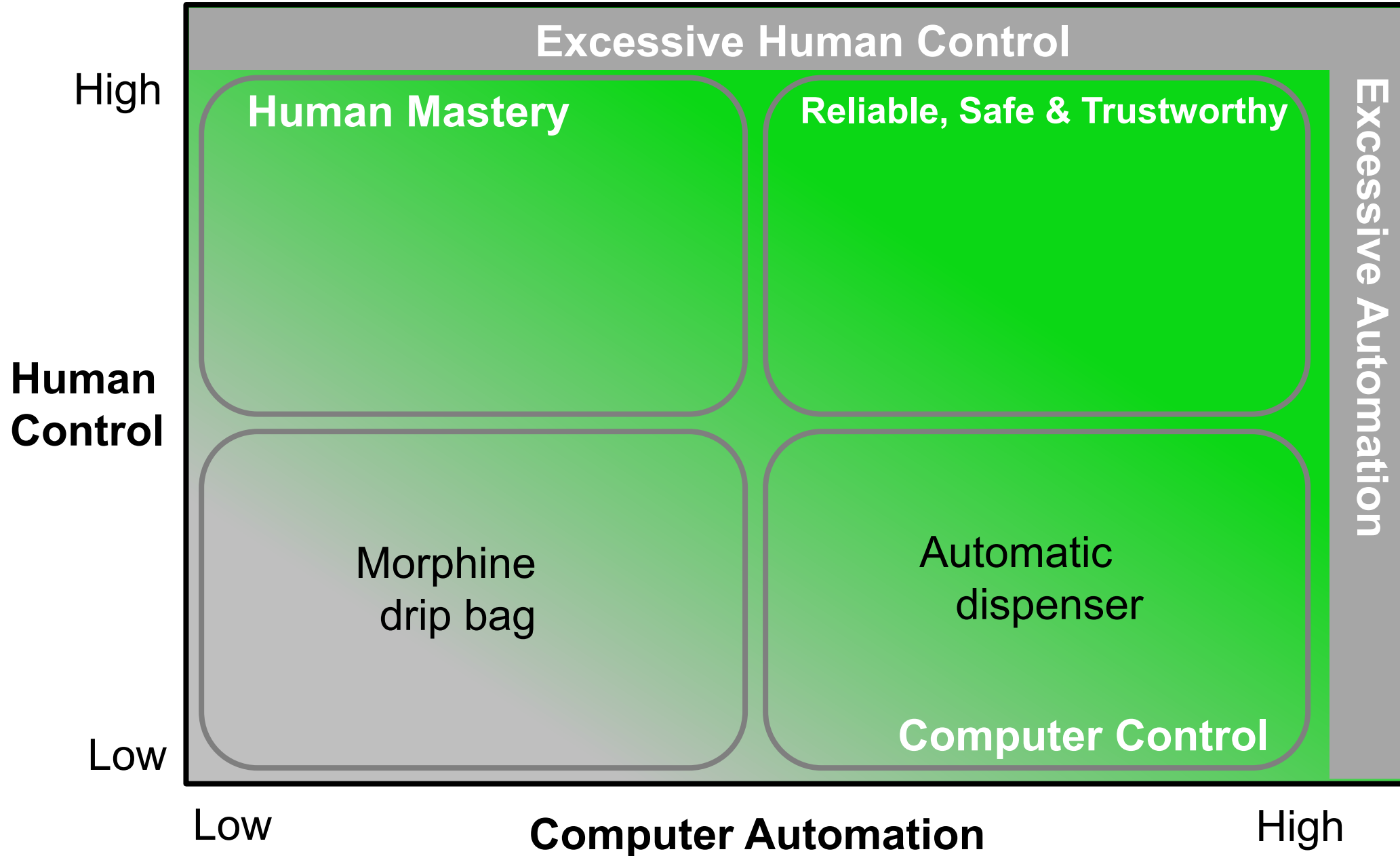
Human-Centered AI



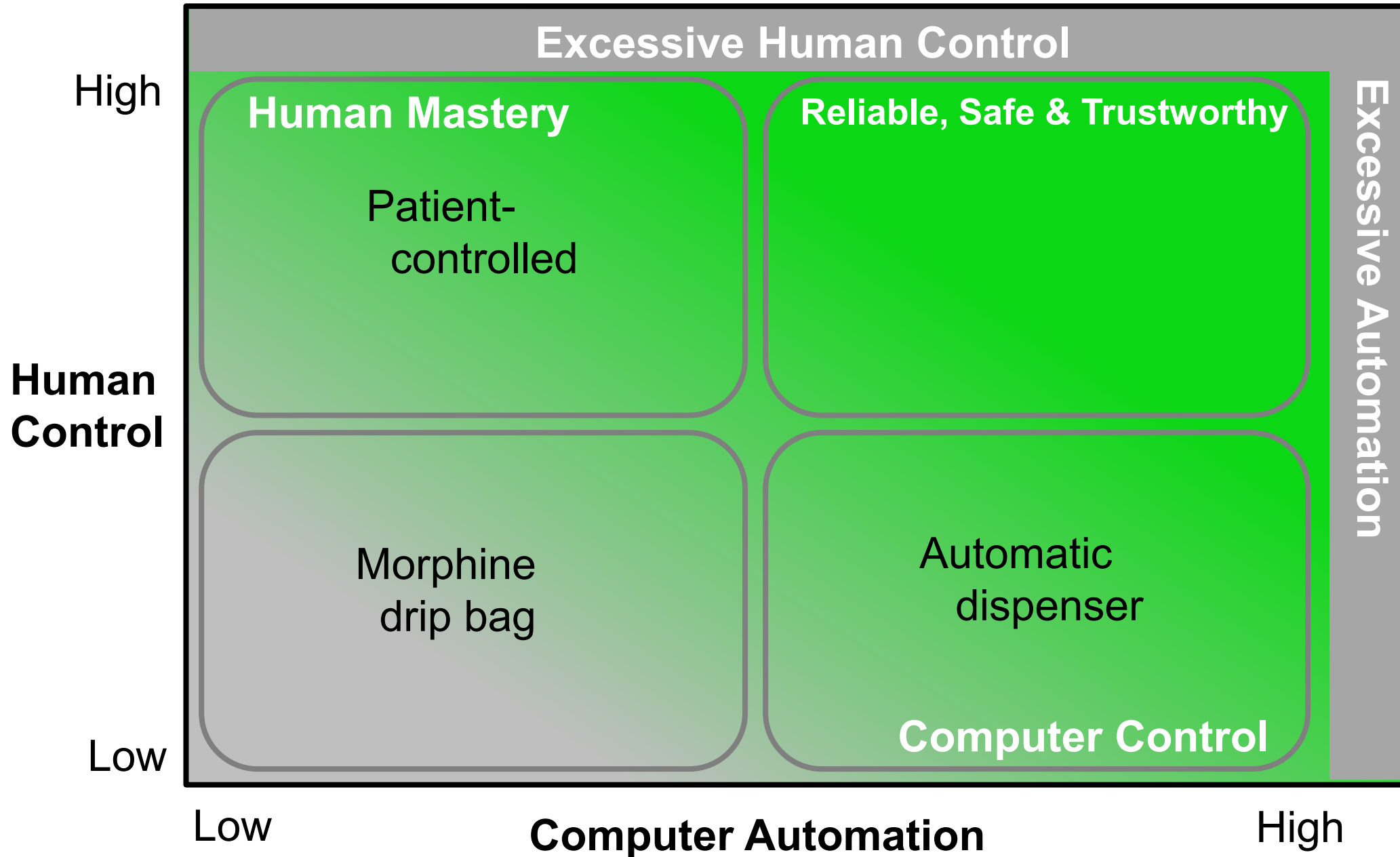
Pain control designs



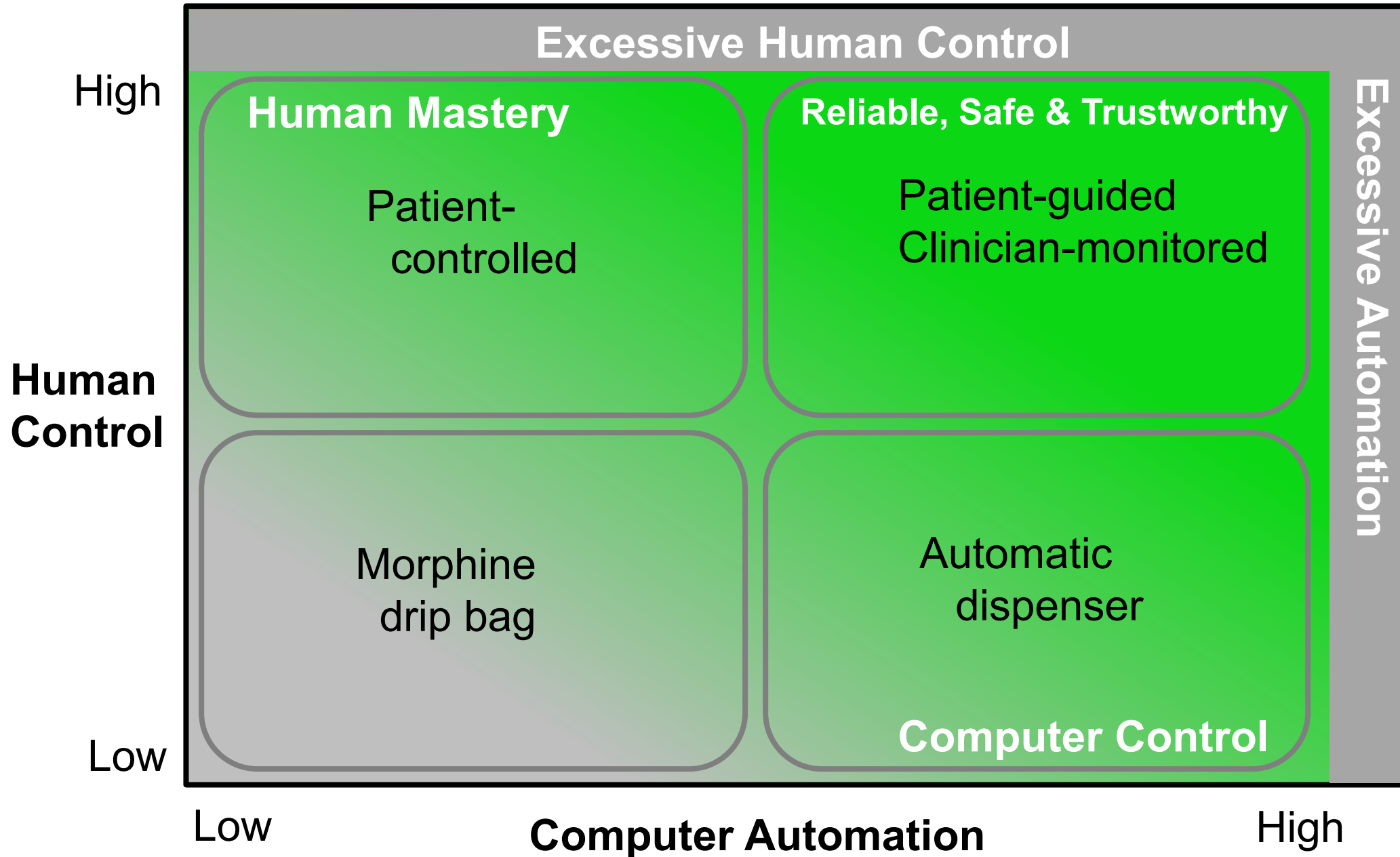
Pain control designs



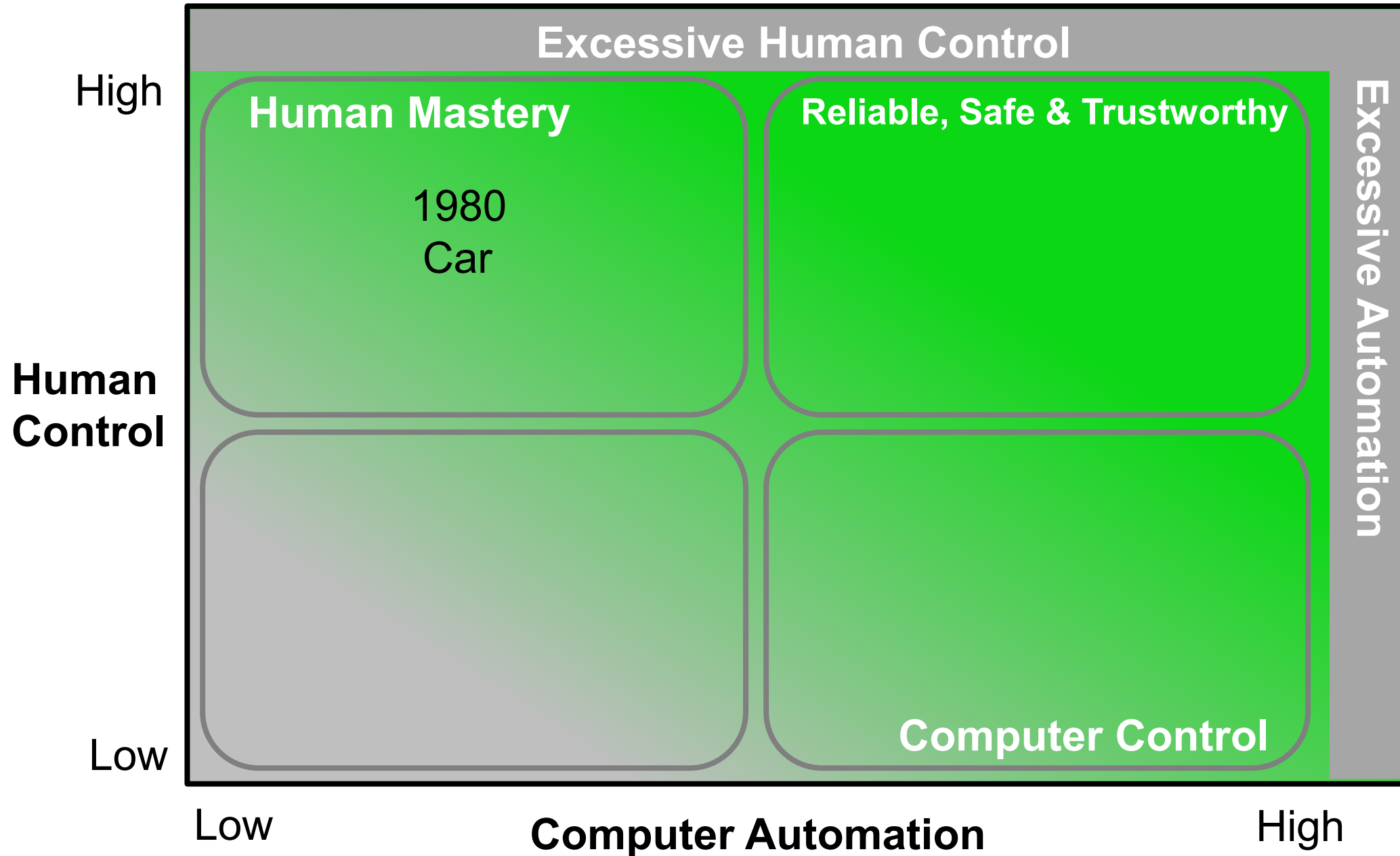
Pain control designs



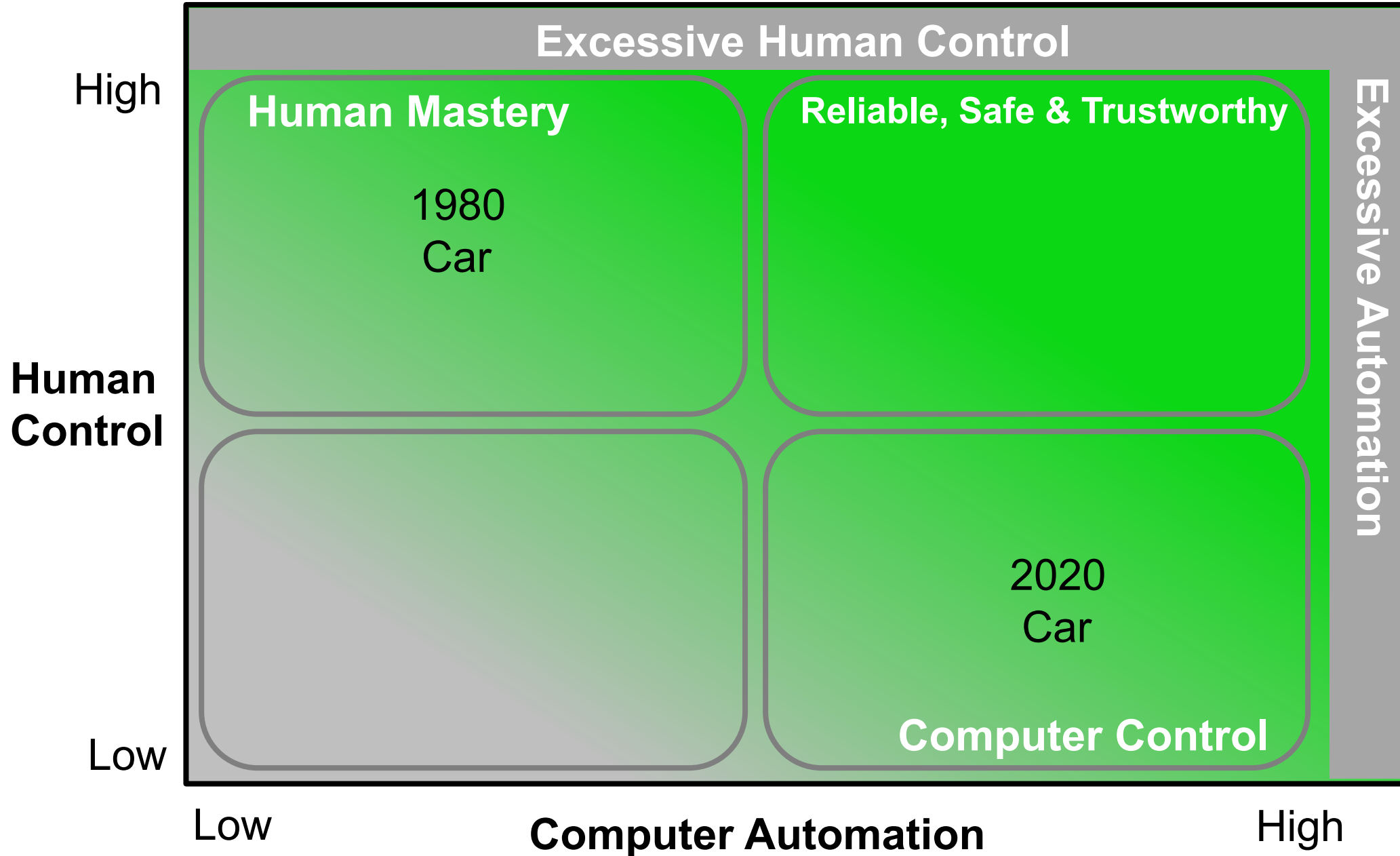
Pain control designs



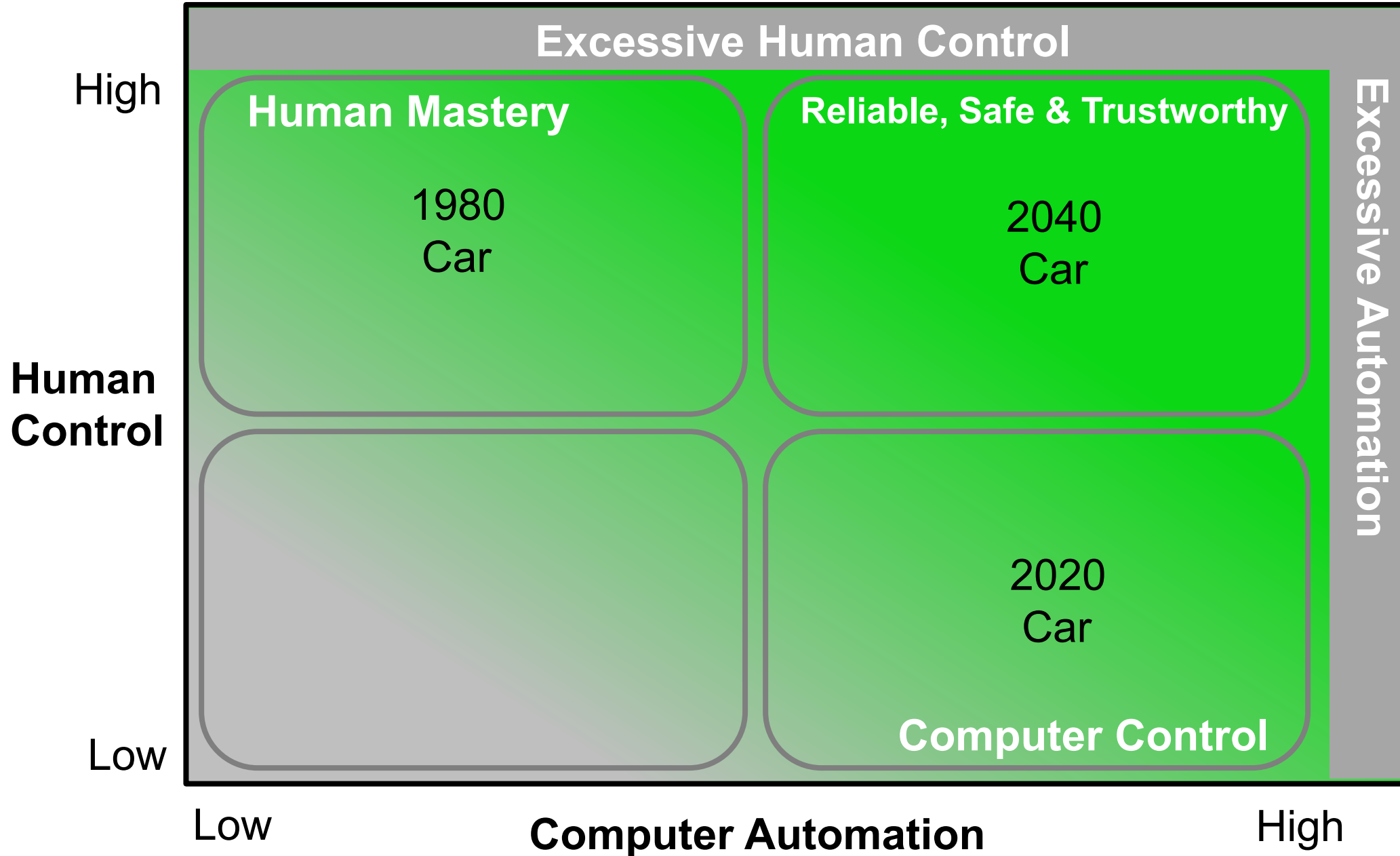
Car control designs



Car control designs



Car control designs



Summary of Design Strategies

Goal: High levels

of human control & computer automation

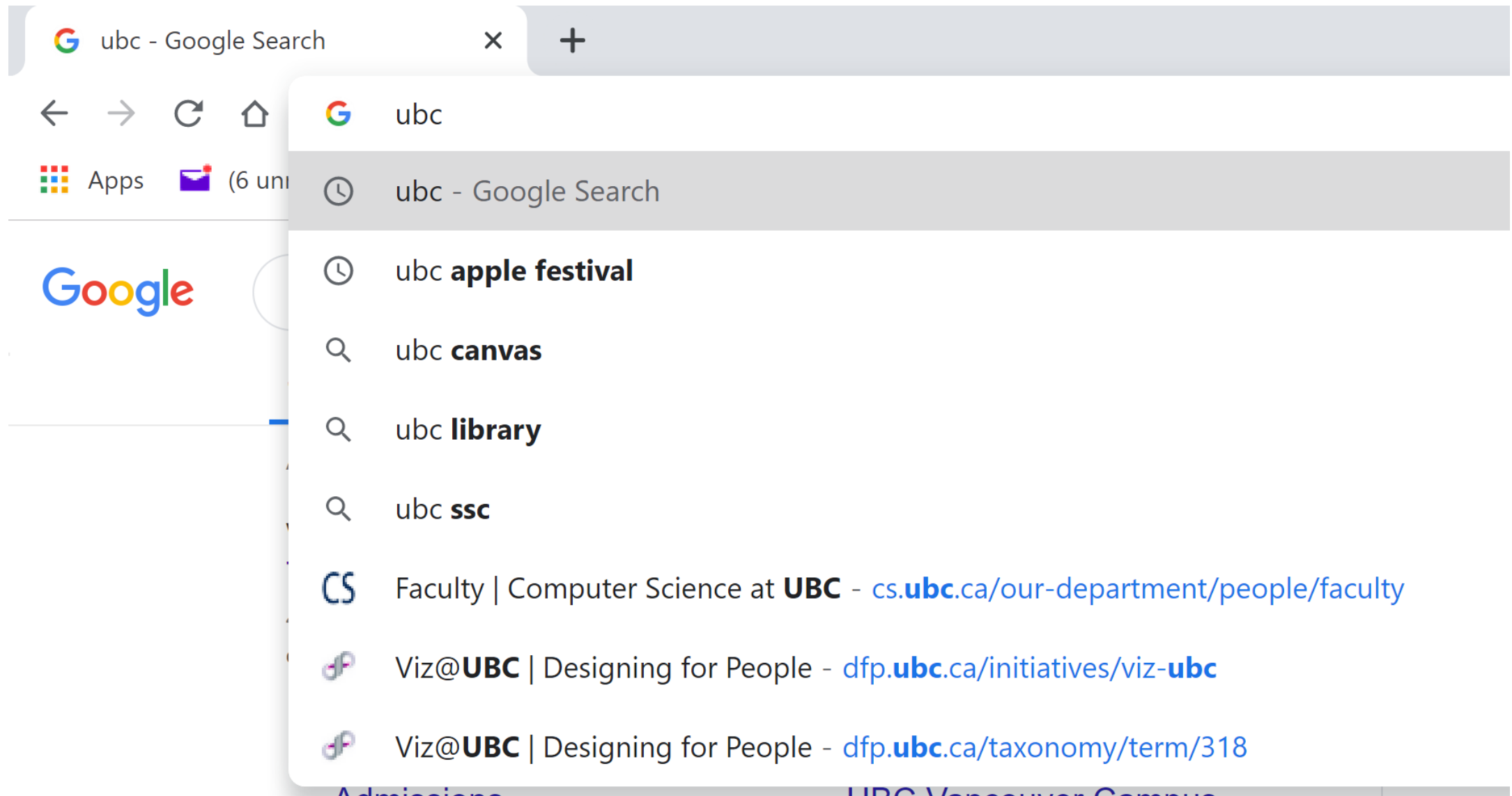
→ RST: Reliable, Safe & Trustworthy

→ Comprehensible, predictable & controllable

→ Continuous display of status

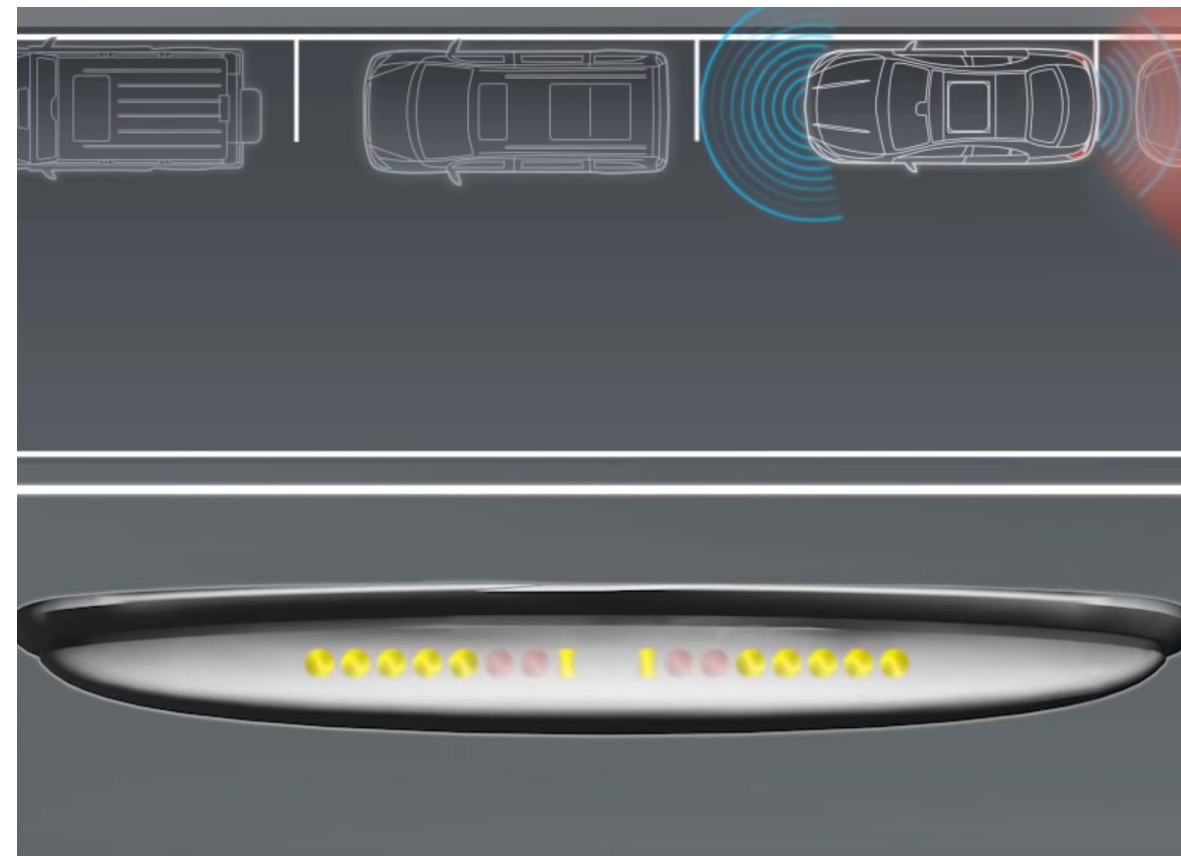
→ Informative feedback

Predictive Search Keeps Users in Control



Mercedes-Benz Parking Assist

Contextually Rich Informative Feedback



<https://youtu.be/NXIMjL39c2Q>

Agency plus automation: Designing artificial intelligence into interactive systems

Jeffrey Heer^{a,1}

^aPaul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA 98195-2350

Edited by Ben Shneiderman, University of Maryland, College Park, MD, and accepted by Editorial Board Member Eva Tardos October 15, 2018 (received for review May 22, 2018)

Much contemporary rhetoric regards the prospects and pitfalls of using artificial intelligence techniques to automate an increasing of such systems may come to overly rely on computational suggestions, leading to a potential loss of critical engagement

to productively employ AI methods
while ensuring that people remain in control:
unconstrained in pursuing complex goals
and exercising domain expertise



RESPONSIBILITY

RESPONSIBILITY

**Accountability, Liability, Fairness, Transparency,
Explainability & Interpretability**

RESPONSIBILITY

**Accountability, Liability, Fairness, Transparency,
Explainability & Interpretability**

Design systems using **reliable** strategies

- ✓ **Audit trails to review failures & near misses**
- ✓ **Benchmark tests for verification & validation**
- ✓ **Continuous improvement & process monitoring**
- ✓ **Data quality & bias testing**
- ✓ **Explainable user interfaces**

RESPONSIBILITY

**Accountability, Liability, Fairness, Transparency,
Explainability & Interpretability**

Seek culture of **safety** by open management strategies:

- ✓ Leadership commitment to safety
- ✓ Open discussion of failures & near misses
- ✓ Internal oversight boards for problems & future plans
- ✓ Public reports of failures & future plans

RESPONSIBILITY

**Accountability, Liability, Fairness, Transparency,
Explainability & Interpretability**

Develop **trustworthy** processes for independent oversight

- ✓ Government regulation
- ✓ Insurance companies require standards
- ✓ Accounting firms audit systems
- ✓ Voluntary industry plans: Partnership on AI
- ✓ Professional societies promote education & ethics

Human-Centered AI

Amplify, Augment, Enhance & Empower People

→ 1000-fold improvements in capabilities

Information

Photography

Search

Navigation

Email & Text

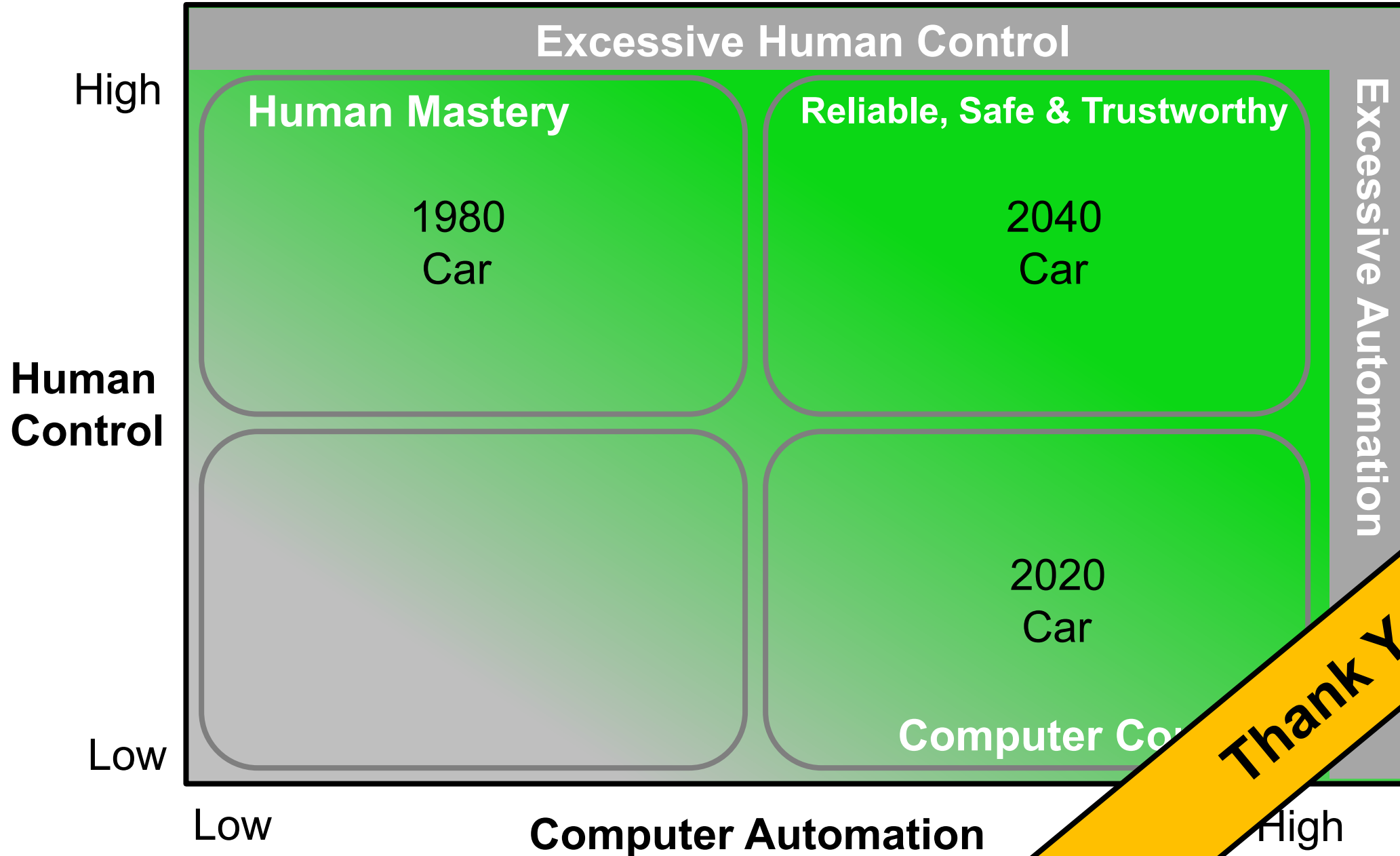
Business Formation

→ RST: Reliable, Safe & Trustworthy

→ More people, more creative, more often

→ Self-efficacy & societal benefits

Car control designs





hcil.umd.edu

HCIL 37th Annual Symposium: May 28, 2020

