NCCI — *The Source You Trust*

Human Factors—

Expanding the Science of Predictive Analytics and Artificial Intelligence (AI)

James Guszcza

Research Affiliate, Center for Advanced Study in the Behavioral Sciences, Stanford University

Human Factors—
Expanding the Science of Predictive Analytics and Artificial Intelligence (AI)

James Guszcza
NCCI AIS
Orlando
May 11, 2022

AI is …

AI is *the new electricity*.

— *Andrew Ng*

Public dialog… too often uses the term AI as *an intellectual wildcard*, one that makes it difficult to reason about the scope and consequences of emerging technology…

This is not the classical case of the public not understanding the scientists—here the scientists are often as befuddled as the public.

— *Michael Jordan, UC Berkeley*
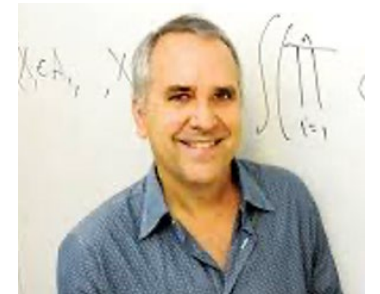
AI is *an ideology*, not a technology.
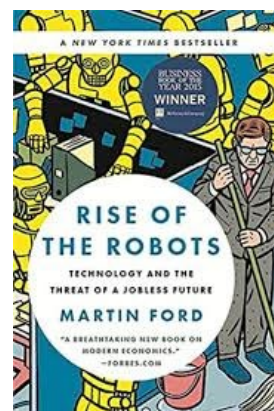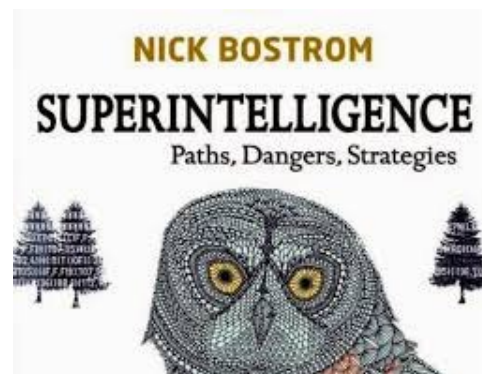
— *Jaron Lanier and Glen Weyl*

The promise of Artificial Intelligence

OpenAI's mission is to ensure that artificial general intelligence (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity.

The company's mission statement was "solve intelligence", and then use it to solve everything else

NICK BOSTROM
SUPERINTELLIGENCE
Paths, Dangers, Strategies

"About 47% of total US employment is at risk [of computerization]."

-- Frey/Osborne (Oxford)

A NEW YORK TIMES BESTSELLER
BUSINESS BOOK OF THE YEAR 2015 WINNER
RISE OF THE ROBOTS
TECHNOLOGY AND THE THREAT OF A JOBLESS FUTURE
MARTIN FORD
"A BREATHTAKING NEW BOOK ON MODERN ECONOMICS."
—FORBES.COM

"We should stop training radiologists now. It's just completely obvious that within five years, deep learning is going to do better than radiologists"

-- Geoffrey Hinton 2016

The prevalence of "Artificial Stupidity"

Through 2020, 80% of AI projects will remain alchemy, run by wizards whose talents will not scale in the organization.

THE VERGE
Amazon reportedly scraps internal AI recruiting tool that was biased against women
The secret program penalized applications that contained the word "women's"
By James Vincent | @jjvincent | Oct 10, 2018, 7:09am EDT

Tesla says driver ignored warnings from Autopilot in fatal California crash

Researchers made an OpenAI GPT-3 medical chatbot as an experiment. It told a mock patient to kill themselves

Racial bias skews algorithms widely used to guide care from heart surgery to birth, study finds

Hundreds of AI tools have been built to catch covid. None of them helped.

TayTweets ✔
@TayandYou
Following

@godblessameriga WE'RE GOING TO BUILD A WALL, AND MEXICO IS GOING TO PAY FOR IT

RETWEETS LIKES
3   5

1:47 AM - 24 Mar 2016

The Hype Machine
How Social Media Disrupts Our Elections, Our Economy, and Our Health – and How We Must Adapt
Sinan Aral

Smart *technologies* are unlikely to engender smart *outcomes*
unless they are designed to promote smart *adoption*
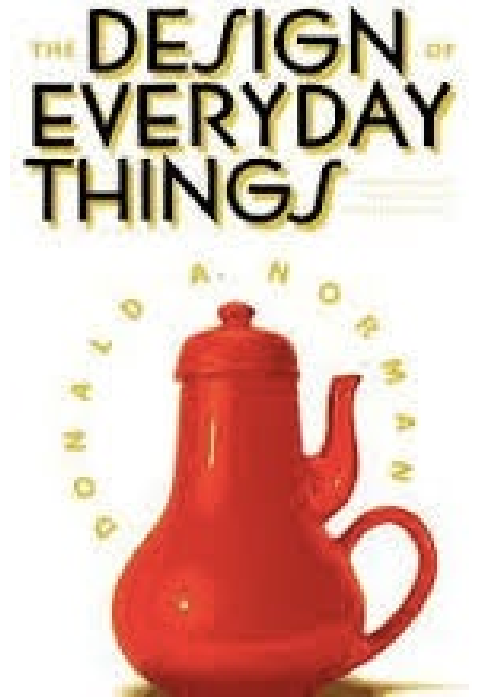on the part of human end users.

→

Effective and Ethical AI needs human-centered design

The AI revolution needs a design revolution

The problem with the designs of most engineers is that they are too logical.

We have to accept human behavior the way it is, not the way we would wish it to be.

— Don Norman, The Design of Everyday Things

Human-centricity: understanding the user

By analogy:
AI technologies will yield better outcomes when they are
designed for the brains of Humans (not "Econs")

THX FOR THE PULLING HANDLES
AND YOUR INSTRUCTIONS TO BE PUSHING THEM

The control room and computer interfaces
at Three Mile Island could not have been
more confusing if they had tried.
— Don Norman

AI and "thinking slow"

THE *NEW YORK TIMES* BESTSELLER

THINKING, FAST AND SLOW

DANIEL KAHNEMAN

WINNER OF THE NOBEL PRIZE IN ECONOMICS

"[A] masterpiece . . . This is one of the greatest and most engaging collections of insights into the human mind I have read." —WILLIAM EASTERLY, *Financial Times*

An Algorithm That Grants Freedom, or Takes It Away

Across the United States and Europe, software is making probation decisions and predicting whether teens will commit crime. Opponents want more human oversight.

Two perspectives on recidivism algorithms

Automatic pilot is an algorithm... We have learned that automatic pilot is more reliable than an individual human pilot.

The same is going to happen here.

— Richard Berk, U. Penn

Does a computer know I might have to go to a doctor's appointment on Friday at 2 o'clock [so cannot visit the probation office]?

How is it going to understand me as it is dictating everything that I have to do?

I can't explain my situation to a computer...

But I can sit here and interact with you, and you can see my expressions and what I am going through.

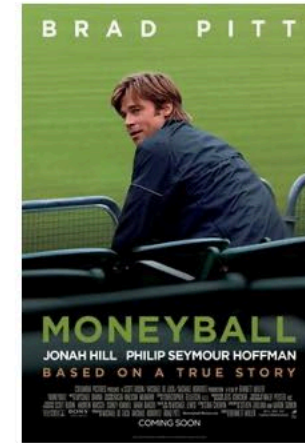*— Darnell Gates, Philadelphia*

Why experts need algorithms

Human judges are not merely worse than optimal regression equations...

They are worse than almost any regression equation.

— Richard Nisbett and Lee Ross

BRAD PITT

MONEYBALL

JONAH HILL   PHILIP SEYMOUR HOFFMAN

BASED ON A TRUE STORY

COMING SOON

Bias

"The places where people are most worried about bias are actually where algorithms have the greatest potential to reduce bias."

– Sendhil Mullainathan

Noise

"We have too much emphasis on bias and not enough emphasis on random noise."

– Daniel Kahneman

# Why algorithms can't (today) replace experts

Judgment requires not only registering the world but doing so in ways appropriate to circumstances.

That is an incredibly high bar.

It requires that a system be oriented toward the world itself, not merely the representations it takes as inputs.

It must be able to distinguish appearance from reality — and defer to reality as the authority.

— Brian Cantwell Smith

THE PROMISE OF ARTIFICIAL INTELLIGENCE

RECKONING AND JUDGMENT

Brian Cantwell Smith
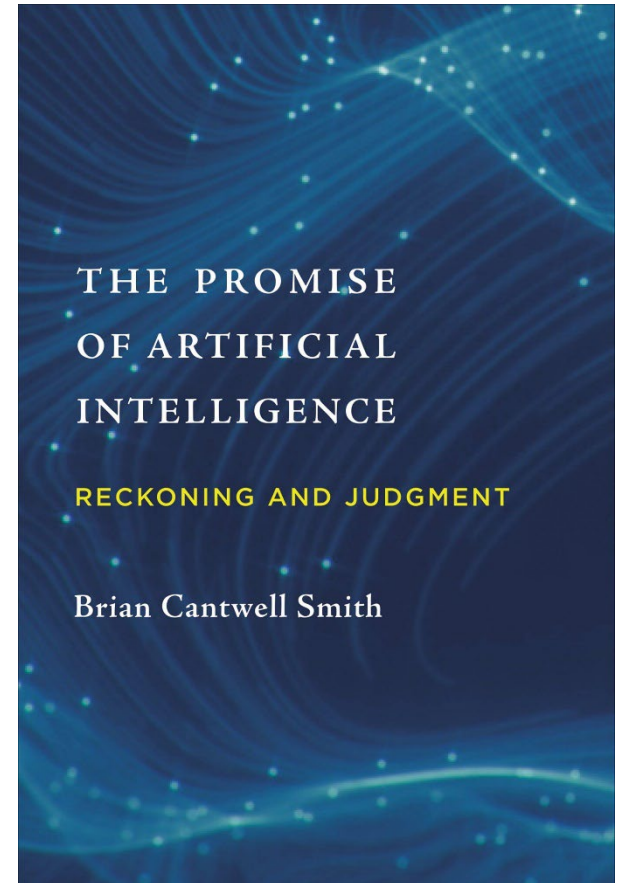
The AI paradox ("The hard problems are easy, and the easy problems are hard.")

Human strengths:
- Strategy
- Causal understanding
- Commonsense reasoning
- Contextual awareness
- Empathy
- Ethical reasoning
- Hypothesis formation
- "Judgment"

Computer strengths:
- Tactics
- Pattern recognition
- Consistency (avoid "noise")
- Rationality   (avoid "bias")
- Brute force
- Narrowly defined, repetitive tasks
- Idiot savant capabilities
- "Reckoning"

WHEN A USER TAKES A PHOTO, THE APP SHOULD CHECK WHETHER THEY'RE IN A NATIONAL PARK...

SURE, EASY GIS LOOKUP. GIMME A FEW HOURS.

...AND CHECK WHETHER THE PHOTO IS OF A BIRD.

I'LL NEED A RESEARCH TEAM AND FIVE YEARS.

IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

<u>Fundamental design implication</u>:
Begin with the assumption of human-machine partnership.
(Machine autonomy should not be the default mode of AI ideation.)

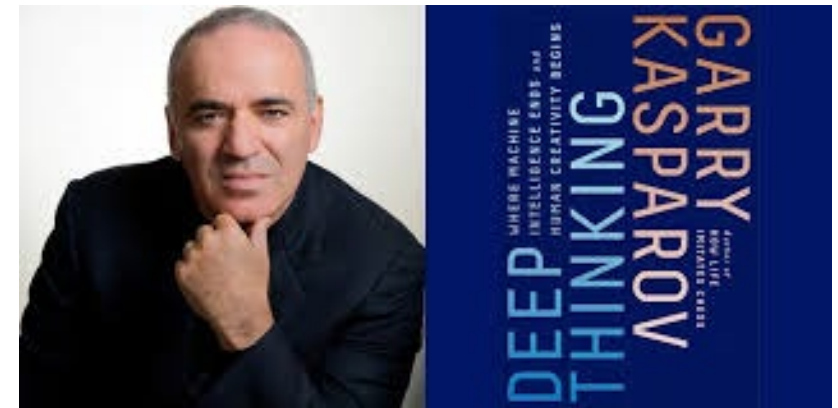# Human-machine hybrid intelligence: A parable

Their skill at manipulating and "coaching" their computers to look very deeply into positions effectively counteracted the superior chess understanding of their grandmaster opponents and the greater computational power of other participants.

Weak human + machine + better process was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process.

— Garry Kasparov, NYRB 2010

And the winners are....: Zack Stephen and Steven Cramton

Designing for human-computer collective intelligence

Weak human + machine + [better process] was superior to a strong computer alone and, more remarkably, superior to a strong human + machine + inferior process.

— Garry Kasparov

Hybrid Intelligence is about more than optimizing algorithms.

It is about "optimizing" processes of **human-machine collaboration**.

Statistics and computer science provides an incomplete scientific framework.

Also needed: Ideas from ethics, psychology, human-centered design, behavioral economics ...

Core principles of hybrid intelligence design

- Design for real-world goals, not machine outputs

- Algorithms aren't enough; they must be embedded in human decision environments

- Decision environments must reflect the *needs*, *behaviors*, and *cognitive capabilities* of the human partner

Implication:
Concepts such as human autonomy, confirmation bias, choice architecture …
should be no less part of hybrid intelligence design than
cross-validation, label bias, data drift …

A needed paradigm shift

A needed paradigm shift
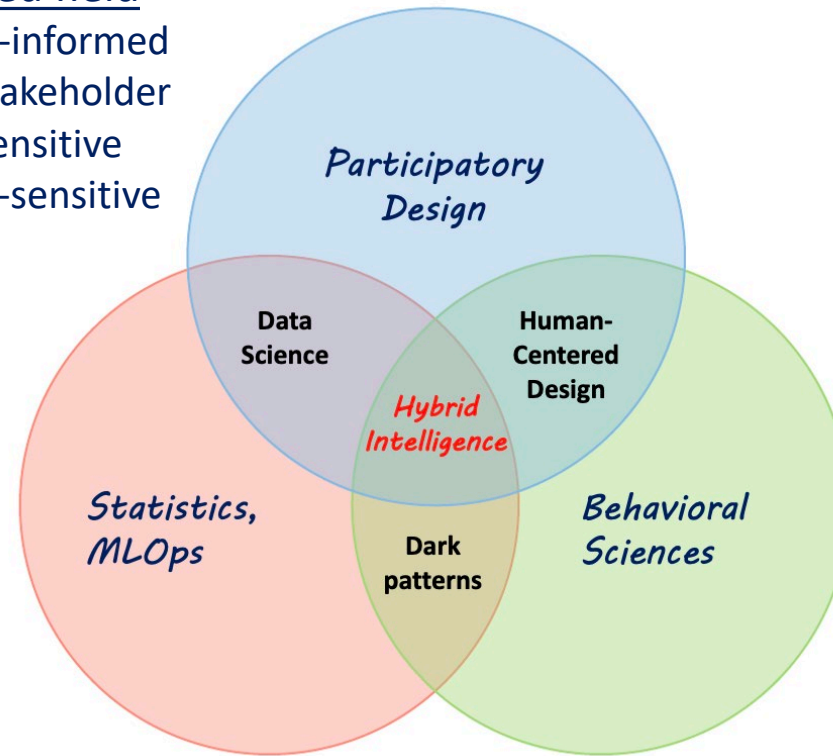

What we often have:
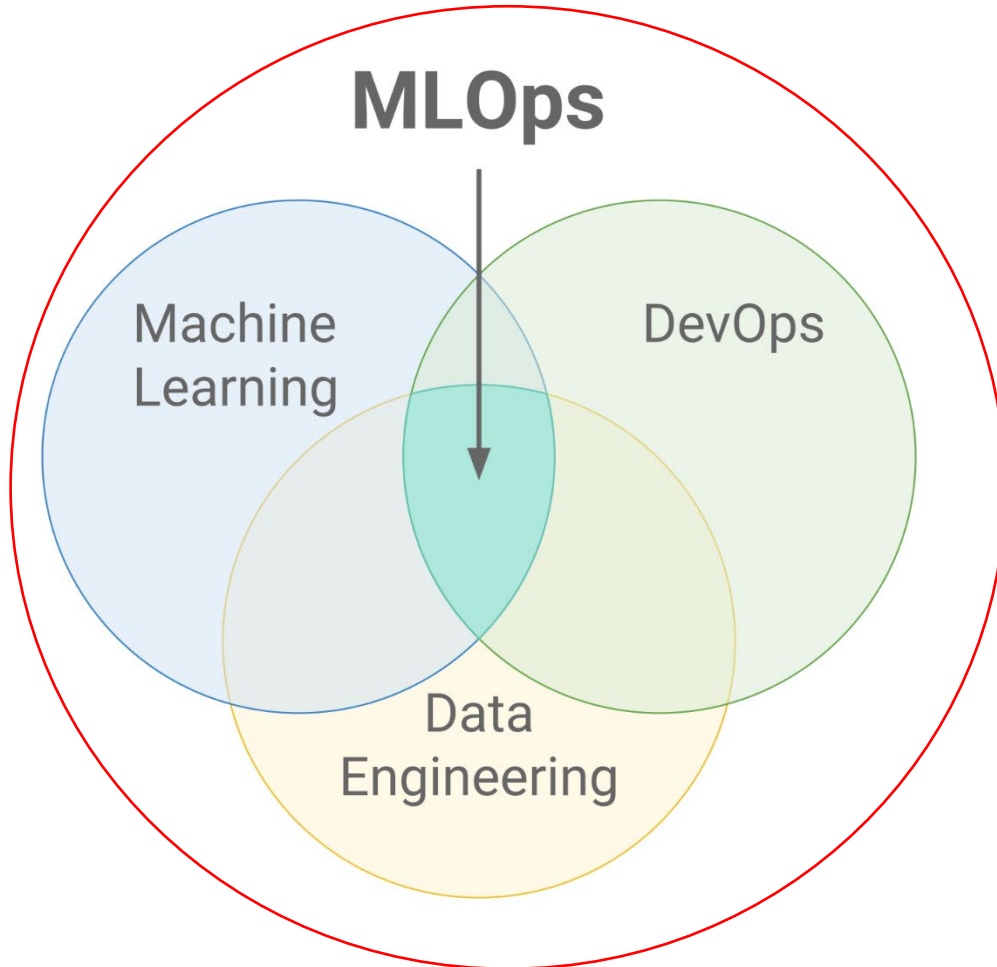
What we typically need:
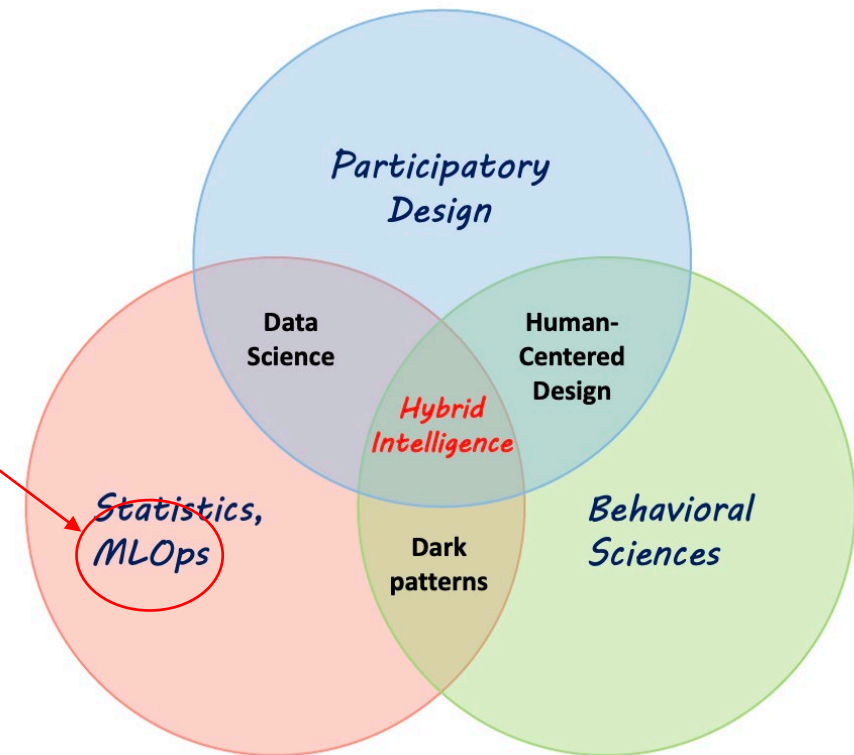
Hybrid intelligence is …

A design-led field
- Domain-informed
- Multi-stakeholder
- Value-sensitive
- Context-sensitive

Grounded in computation AND statistics
- Does more than extract patterns from data
- Also considers how well the data reflects the real world

Grounded in the behavioral sciences
- Behavioral economics
- Organizational design
- Cognitive psychology
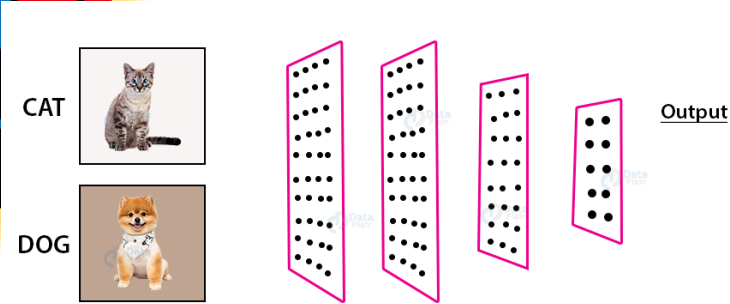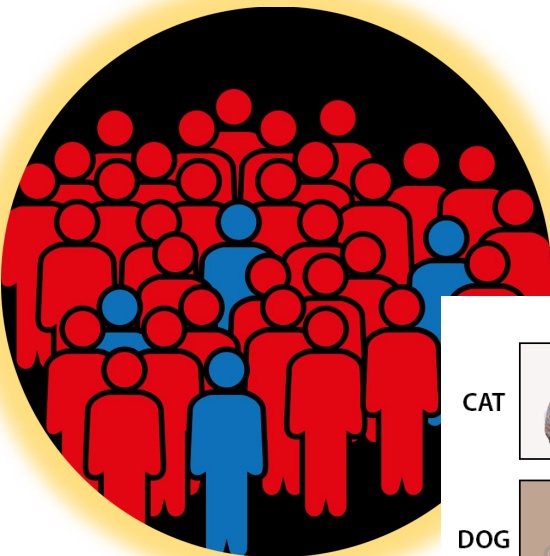- Affective science
- …

Ethics and the need for "greater AI"

Naïve view

"Nudge" view