ISDC 2025
August 6, 2025

How AI Can Help System Dynamics (SD) - and Why AI Needs SD

Donald Martin, Jr.
Global Institute for the Learning Society

Why AI Needs SD

Are we living in a Sci-Fi TV Drama?

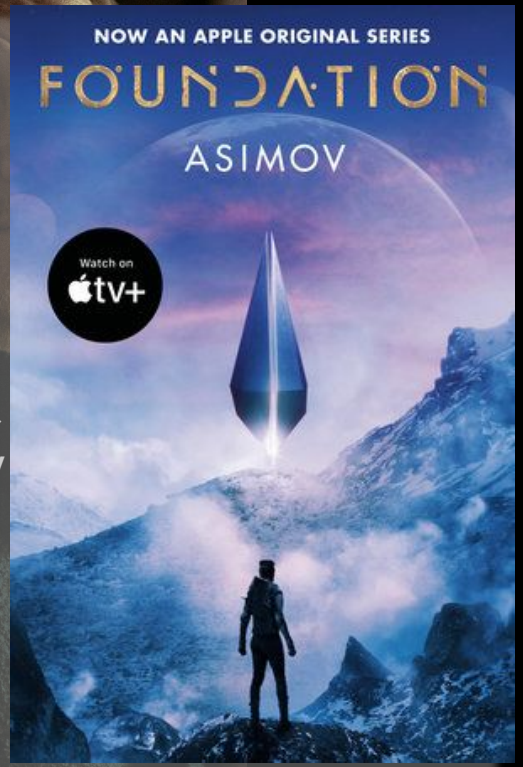
On GPS: AI 'godfather' warns of threat to humanity

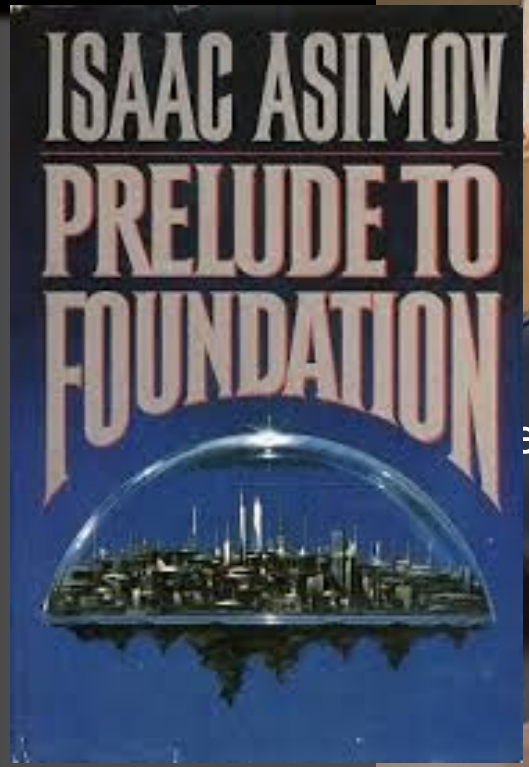
Elon Musk: 10 billion humanoid robots by 2040 at $20K-$25K each

By **Reuters**

October 29, 2024 11:26 AM EDT · Updated October 29, 2024

Hari Seldon &
Psychohistory





ezel / Daneel

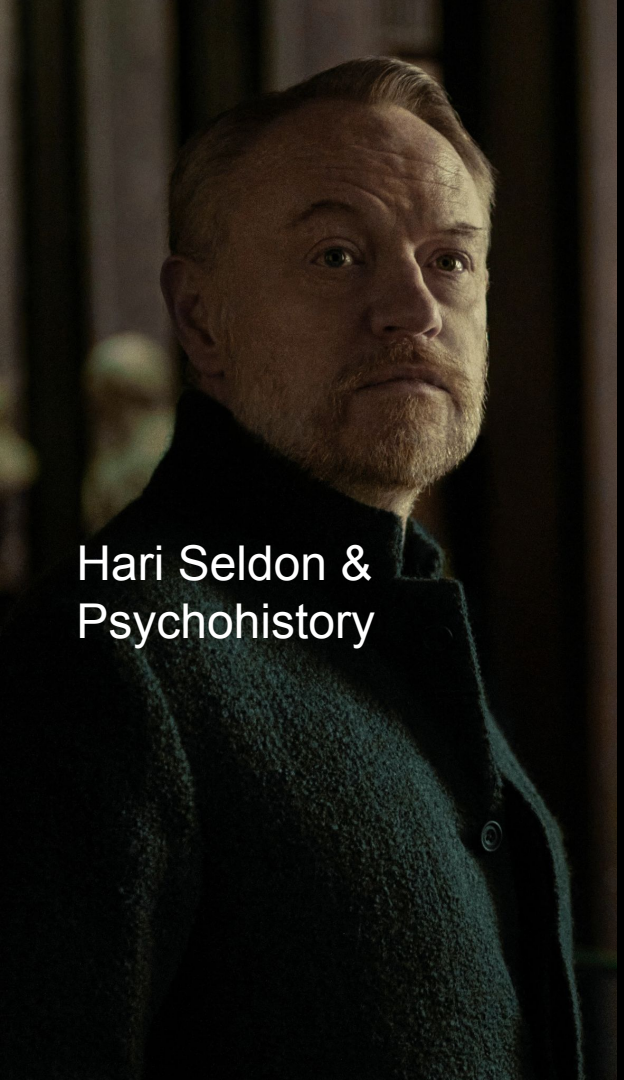
Isaac Asimov's Laws of Robotics (aka AI)

1) a robot may not injure a human being or, through inaction, allow a human being to come to harm;
2) a robot must obey the orders given it by human beings except where such orders would conflict with the First Law;
3) a robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Isaac Asimov's Laws of Robotics (aka AI)

0) A robot may not harm humanity, or, by inaction, allow humanity to come to harm.

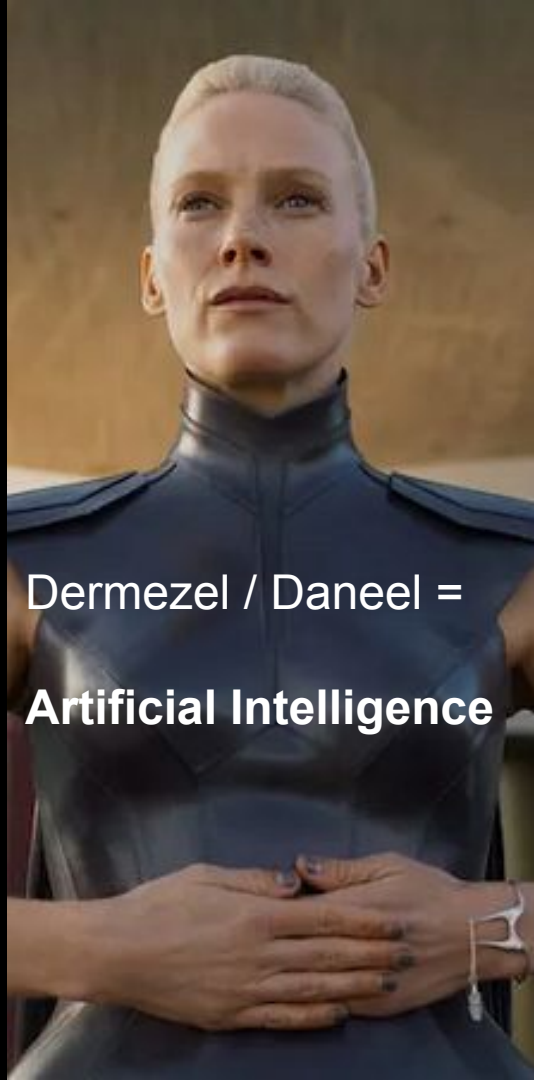
1) a robot may not injure a human being or, through inaction, allow a human being to come to harm except where that would conflict with the Zeroth Law;
2) a robot must obey the orders given it by human beings except where such orders would conflict with the Zeroth or First Law;
3) a robot must protect its own existence as long as such protection does not conflict with the Zeroth, First or Second Law.



Hari Seldon &
Psychohistory

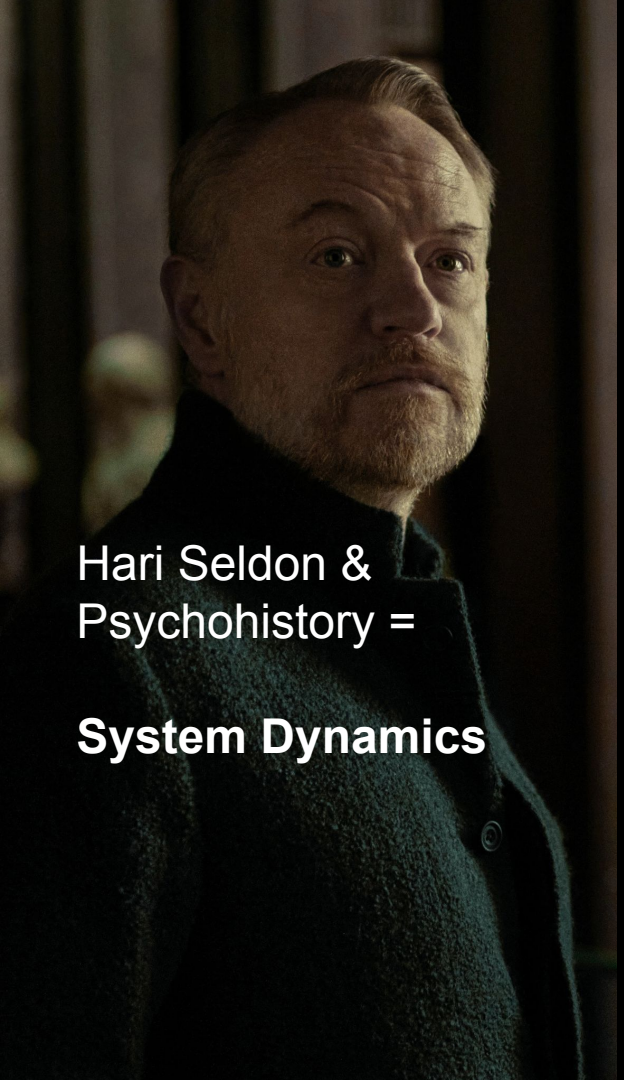
Dermezel:

“But what is humanity? To what can we point when we speak of humanity? **And how can we define harm to humanity?** When will a course of action do more good than harm to humanity as a whole and how can one tell?”



Dermezel / Daneel =

Artificial Intelligence

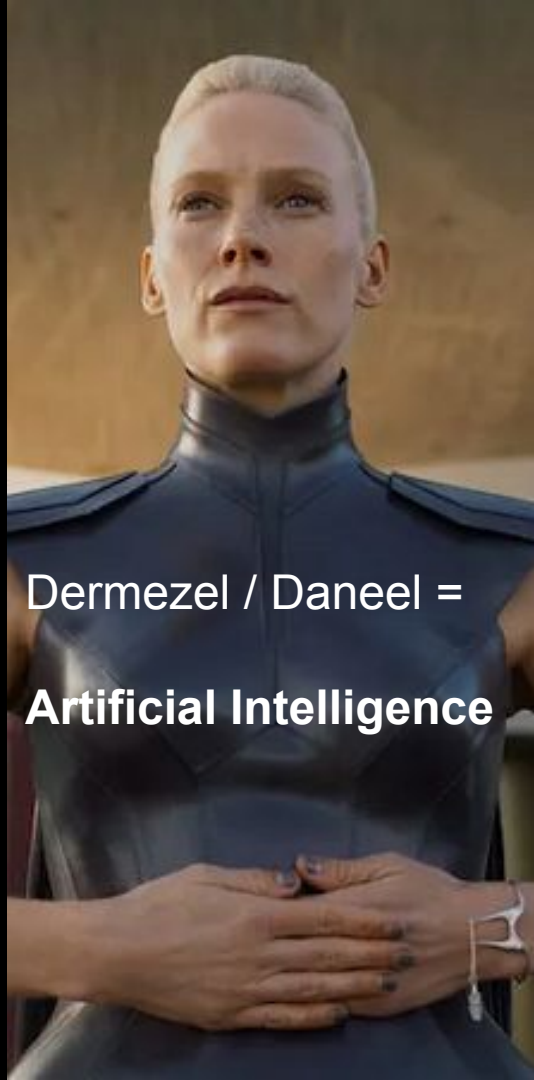


Hari Seldon & Psychohistory =

System Dynamics

<u>Dermezel:</u>

“I realized at once that in psychohistory there was a tool that might make it possible to identify what was good and bad for humanity. ***With it, the decisions we would make would be less blind.”***



Dermezel / Daneel =

Artificial Intelligence

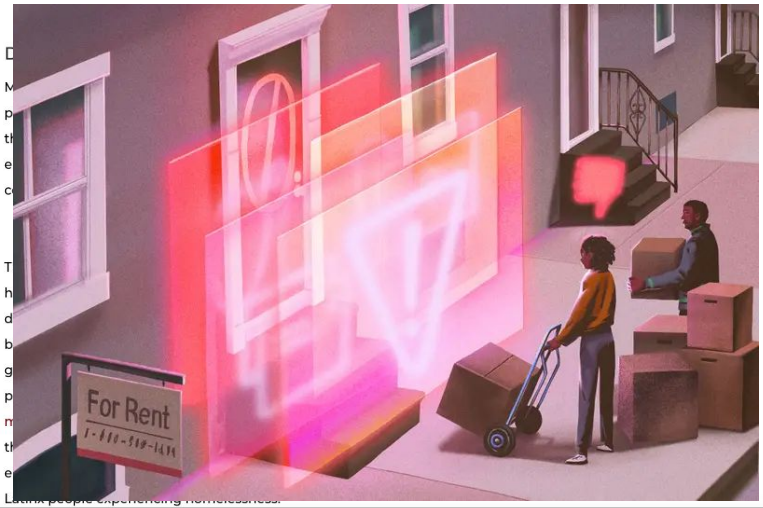
Clear evidence that AI blindness to humanity **is real and harmful** …

THE UNIVERSITY OF CHICAGO
KREISMAN INITIATIVE FOR HOUSING LAW & POLICY
Mansueto Institute for Urban Innovation

About · Opportunities · Symposium · Research · News and Events

AI is Making Housing Discrimination Easier Than Ever Before

Feb 12, 2024 | 0 comments



The Washington Post

Health

Racial bias in a medical algorithm favors white patients over sicker black patients

Scientists discovered racial bias in a widely used medical algorithm that predicts which patients will have complex health needs. (iStock)

By Carolyn Y. Johnson

Oct. 24, 2019 at 11:00 a.m. PDT

Poor problem understanding & formulation were the root cause of this failure.

Obermeyer, Ziad, et al. "Dissecting racial bias in an algorithm used to manage the health of populations." Science 366.6464 (2019): 447-453.

The Washington Post
Democracy Dies in Darkness

Health

Racial bias in a medical algorithm favors white patients over sicker black patients

Scientists discovered racial bias in a widely used medical algorithm that predicts which patients will have complex health needs. (iStock)

By **Carolyn Y. Johnson**

Oct. 24, 2019 at 11:00 a.m. PDT

People not selected for the special programs suffered from nearly 50,000 more chronic diseases than those selected.

Obermeyer, Ziad, et al. "Dissecting racial bias in an algorithm used to manage the health of populations." Science 366.6464 (2019): 447-453.

Slide derived from: Donald Martin, Jr. , Aug 30, 2023, “Epistemic Uncertainty, the AI Problem Understanding Gap and the Necessity of Structured Societal Context Knowledge for Safe, Robust AI”, Vimeo, https://vimeo.com/859590420?fl=pl&fe=vl

How was problem understanding poor?

These critical factors were missing from the conception of the problem.

The false assumption that this was true for all populations led to a racially biased AI system.

Structural Inequities (aka othering)

Increases (+)

- Underdiagnosis — Decreases (-)
- Lack of trust in health care system — Decreases (-)
- The racial wealth gap — Decreases (-)
- Lack of access to affordable health care — Decreases (-)

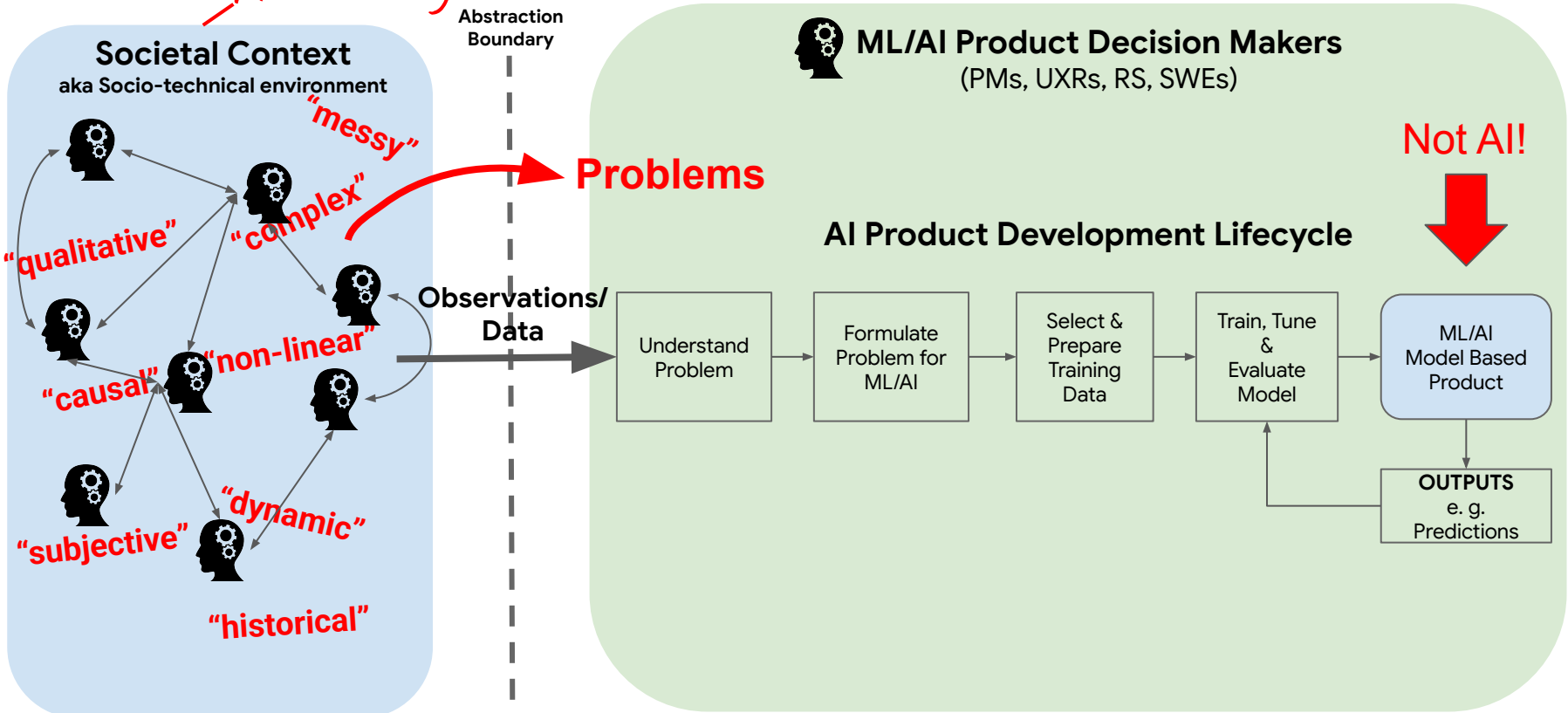
Spending on healthcare ← Increases (+) ← Complex healthcare needs

Proxy

Increases (+)

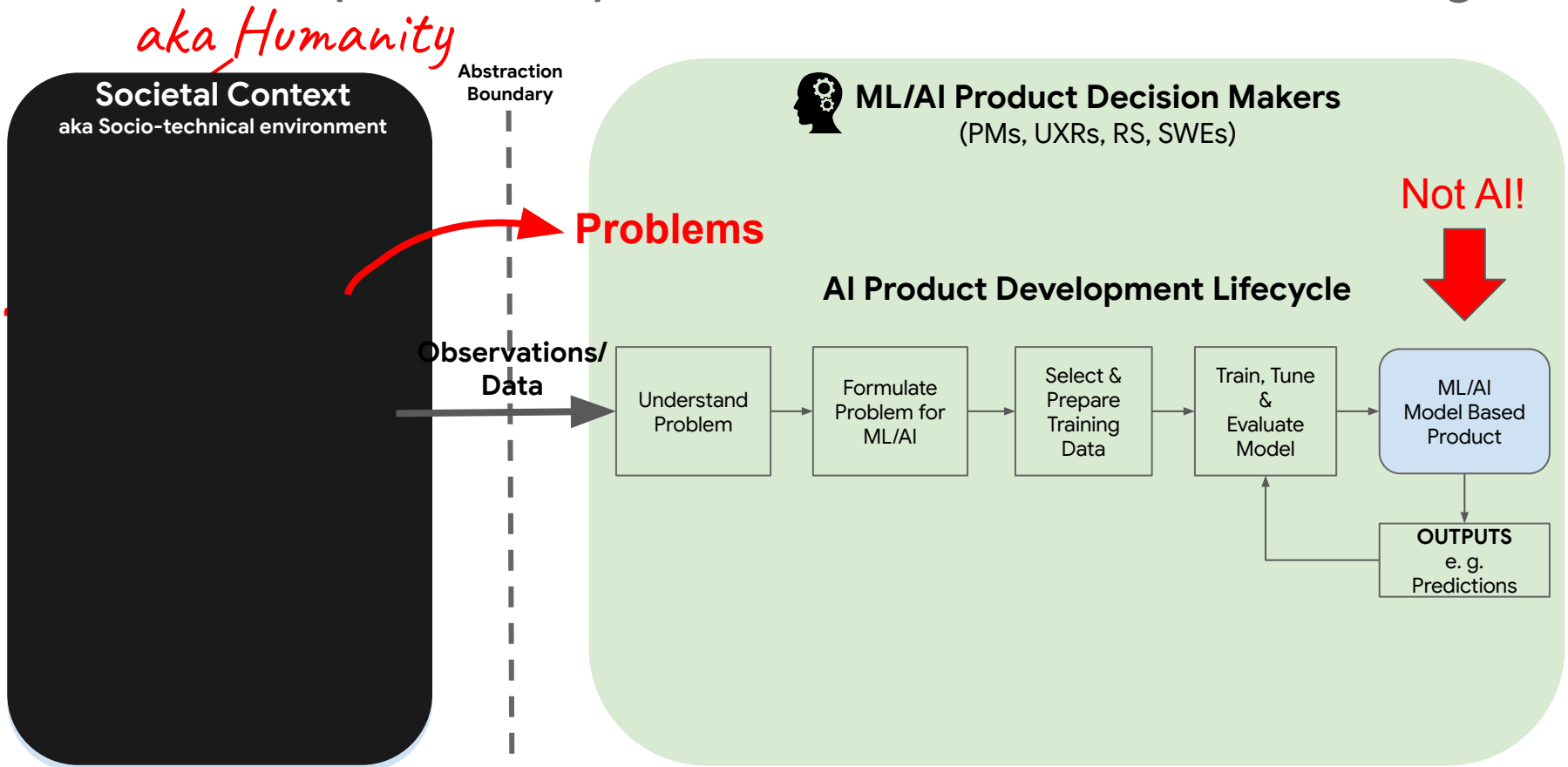
AI is a complex ecosystem of human decision making

aka Humanity



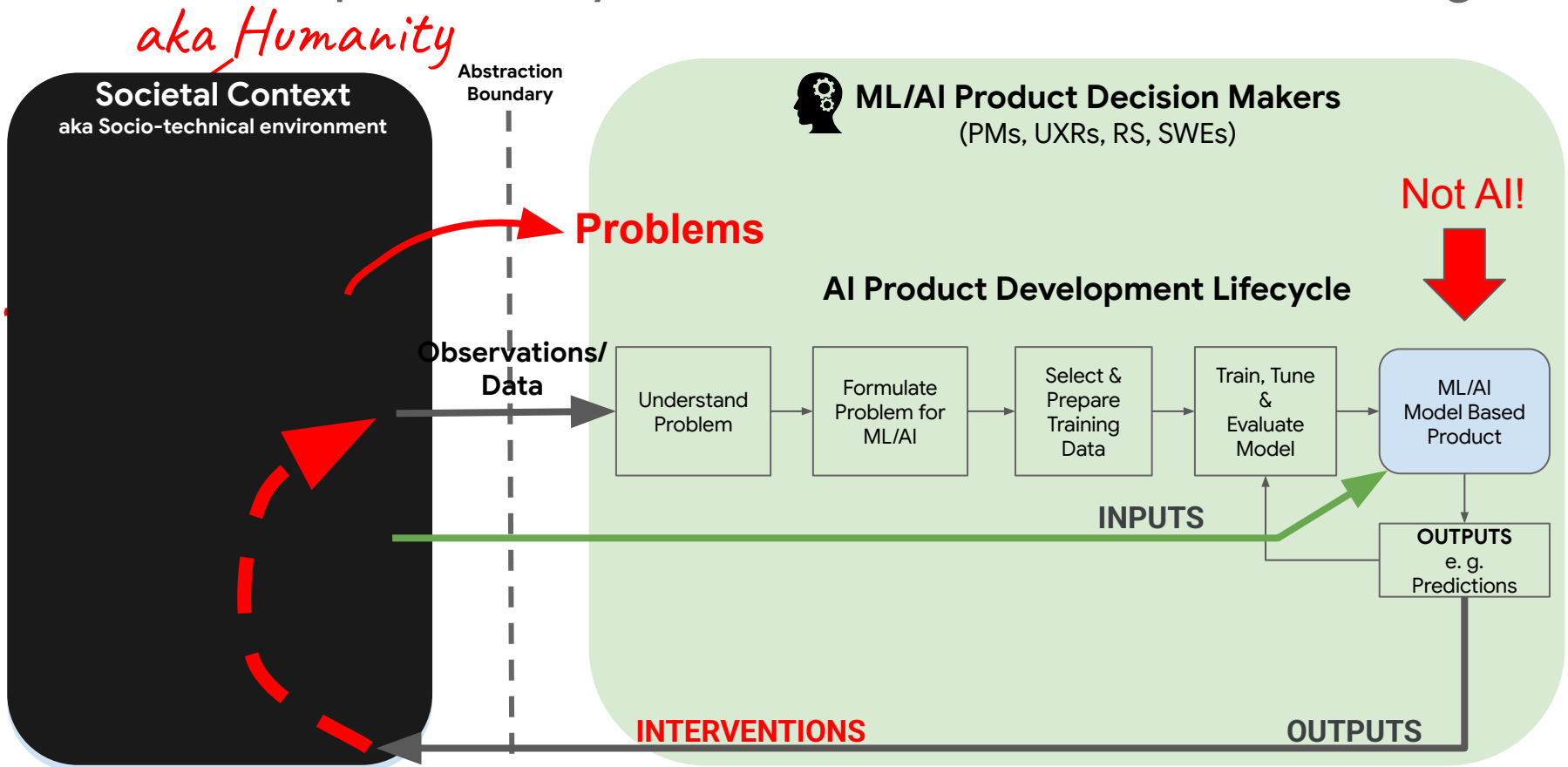
Slide derived from: Donald Martin, Jr. , Aug 30, 2023, "Epistemic Uncertainty, the AI Problem Understanding Gap and the Necessity of Structured Societal Context Knowledge for Safe, Robust AI", Vimeo, https://vimeo.com/859590420?fl=pl&fe=vl

AI is a complex ecosystem of human decision making



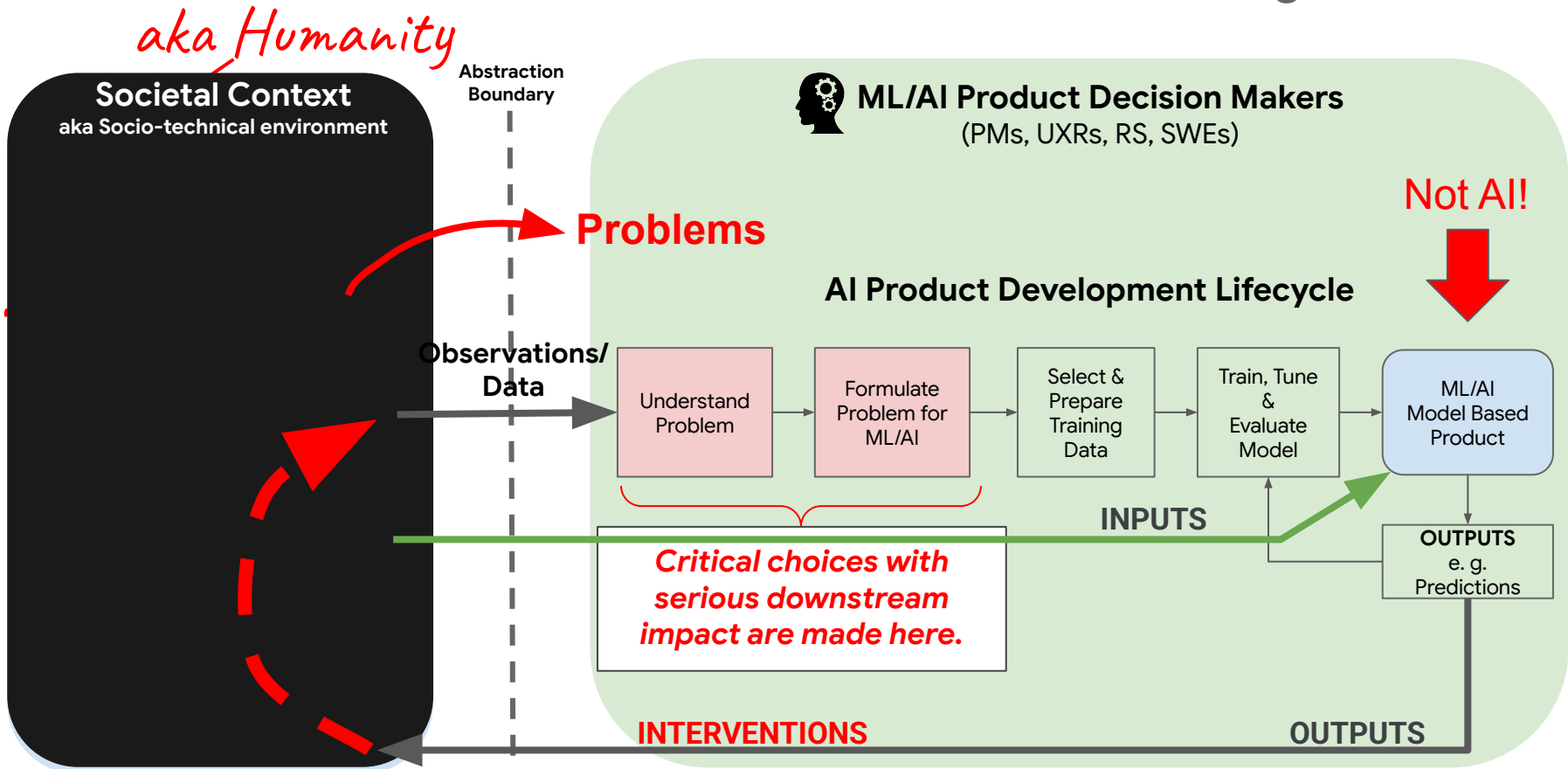
Slide derived from: Donald Martin, Jr. , Aug 30, 2023, “Epistemic Uncertainty, the AI Problem Understanding Gap and the Necessity of Structured Societal Context Knowledge for Safe, Robust AI”, Vimeo, https://vimeo.com/859590420?fl=pl&fe=vl

AI is a complex ecosystem of human decision making



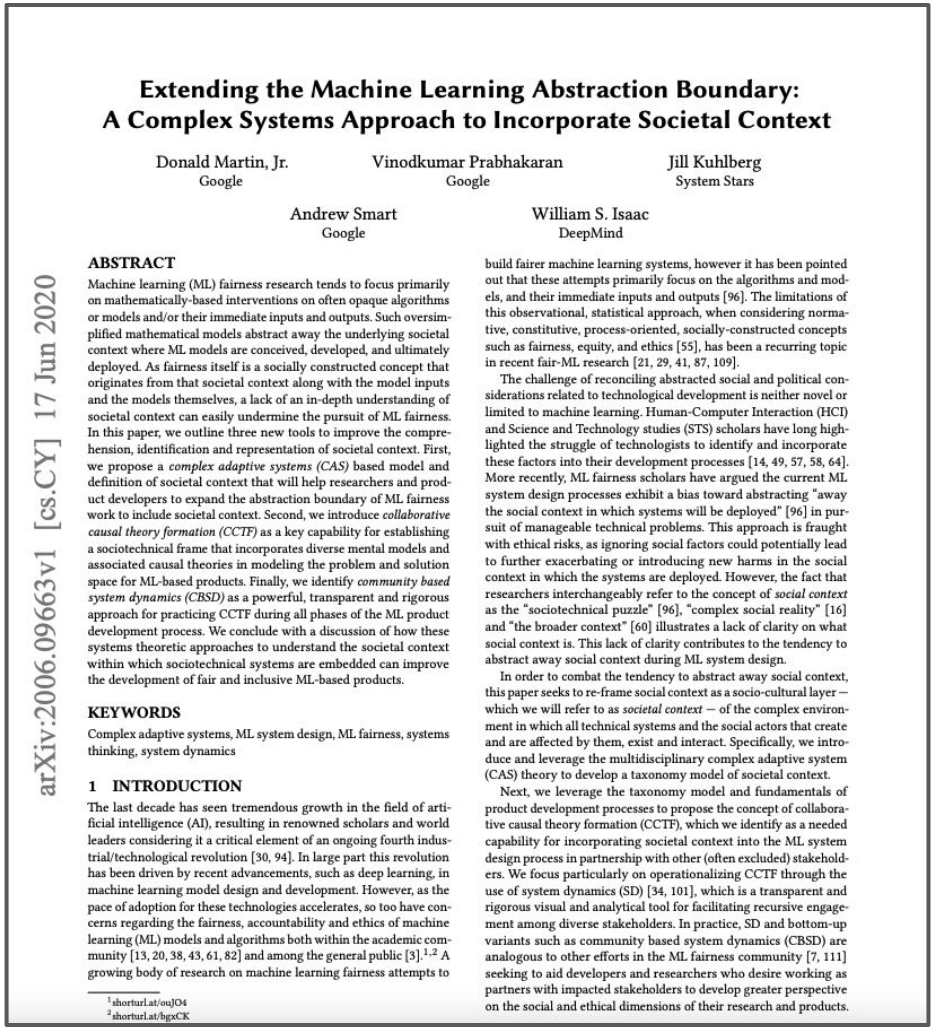
Slide derived from: Donald Martin, Jr. , Aug 30, 2023, “Epistemic Uncertainty, the AI Problem Understanding Gap and the Necessity of Structured Societal Context Knowledge for Safe, Robust AI”, Vimeo, https://vimeo.com/859590420?fl=pl&fe=vl

Problem formulation decisions and choices are high stakes



Slide derived from: Donald Martin, Jr. , Aug 30, 2023, “Epistemic Uncertainty, the AI Problem Understanding Gap and the Necessity of Structured Societal Context Knowledge for Safe, Robust AI”, Vimeo, https://vimeo.com/859590420?fl=pl&fe=vl

To understand problems we must understand societal context.



Extending the Machine Learning Abstraction Boundary: A Complex Systems Approach to Incorporate Societal Context

Donald Martin, Jr.
Google

Vinodkumar Prabhakaran
Google

Jill Kuhlberg
System Stars

Andrew Smart
Google

William S. Isaac
DeepMind

ABSTRACT

Machine learning (ML) fairness research tends to focus primarily on mathematically-based interventions on often opaque algorithms or models and/or their immediate inputs and outputs. Such oversimplified mathematical models abstract away the underlying societal context where ML models are conceived, developed, and ultimately deployed. As fairness itself is a socially constructed concept that originates from that societal context along with the model inputs and the models themselves, a lack of an in-depth understanding of societal context can easily undermine the pursuit of ML fairness. In this paper, we outline three new tools to improve the comprehension, identification and representation of societal context. First, we propose a *complex adaptive systems (CAS)* based model and definition of societal context that will help researchers and product developers to expand the abstraction boundary of ML fairness work to include societal context. Second, we introduce *collaborative causal theory formation (CCTF)* as a key capability for establishing a sociotechnical frame that incorporates diverse mental models and associated causal theories in modeling the problem and solution space for ML-based products. Finally, we identify *community based system dynamics (CBSD)* as a powerful, transparent and rigorous approach for practicing CCTF during all phases of the ML product development process. We conclude with a discussion of how these systems theoretic approaches to understand the societal context within which sociotechnical systems are embedded can improve the development of fair and inclusive ML-based products.

KEYWORDS

Complex adaptive systems, ML system design, ML fairness, systems thinking, system dynamics

1 INTRODUCTION

The last decade has seen tremendous growth in the field of artificial intelligence (AI), resulting in renowned scholars and world leaders considering it a critical element of an ongoing fourth industrial/technological revolution [30, 94]. In large part this revolution has been driven by recent advancements, such as deep learning, in machine learning model design and development. However, as the pace of adoption for these technologies accelerates, so too have concerns regarding the fairness, accountability and ethics of machine learning (ML) models and algorithms both within the academic community [13, 20, 38, 43, 61, 82] and among the general public [3].[1,2] A growing body of research on machine learning fairness attempts to build fairer machine learning systems, however it has been pointed out that these attempts primarily focus on the algorithms and models, and their immediate inputs and outputs [96]. The limitations of this observational, statistical approach, when considering normative, constitutive, process-oriented, socially-constructed concepts such as fairness, equity, and ethics [55], has been a recurring topic in recent fair-ML research [21, 29, 41, 87, 109].

The challenge of reconciling abstracted social and political considerations related to technological development is neither novel or limited to machine learning. Human-Computer Interaction (HCI) and Science and Technology studies (STS) scholars have long highlighted the struggle of technologists to identify and incorporate these factors into their development processes [14, 49, 57, 58, 64]. More recently, ML fairness scholars have argued the current ML system design processes exhibit a bias toward abstracting "away the social context in which systems will be deployed" [96] in pursuit of manageable technical problems. This approach is fraught with ethical risks, as ignoring social factors could potentially lead to further exacerbating or introducing new harms in the social context in which the systems are deployed. However, the fact that researchers interchangeably refer to the concept of *social context* as the "sociotechnical puzzle" [96], "complex social reality" [16] and "the broader context" [60] illustrates a lack of clarity on what social context is. This lack of clarity contributes to the tendency to abstract away social context during ML system design.

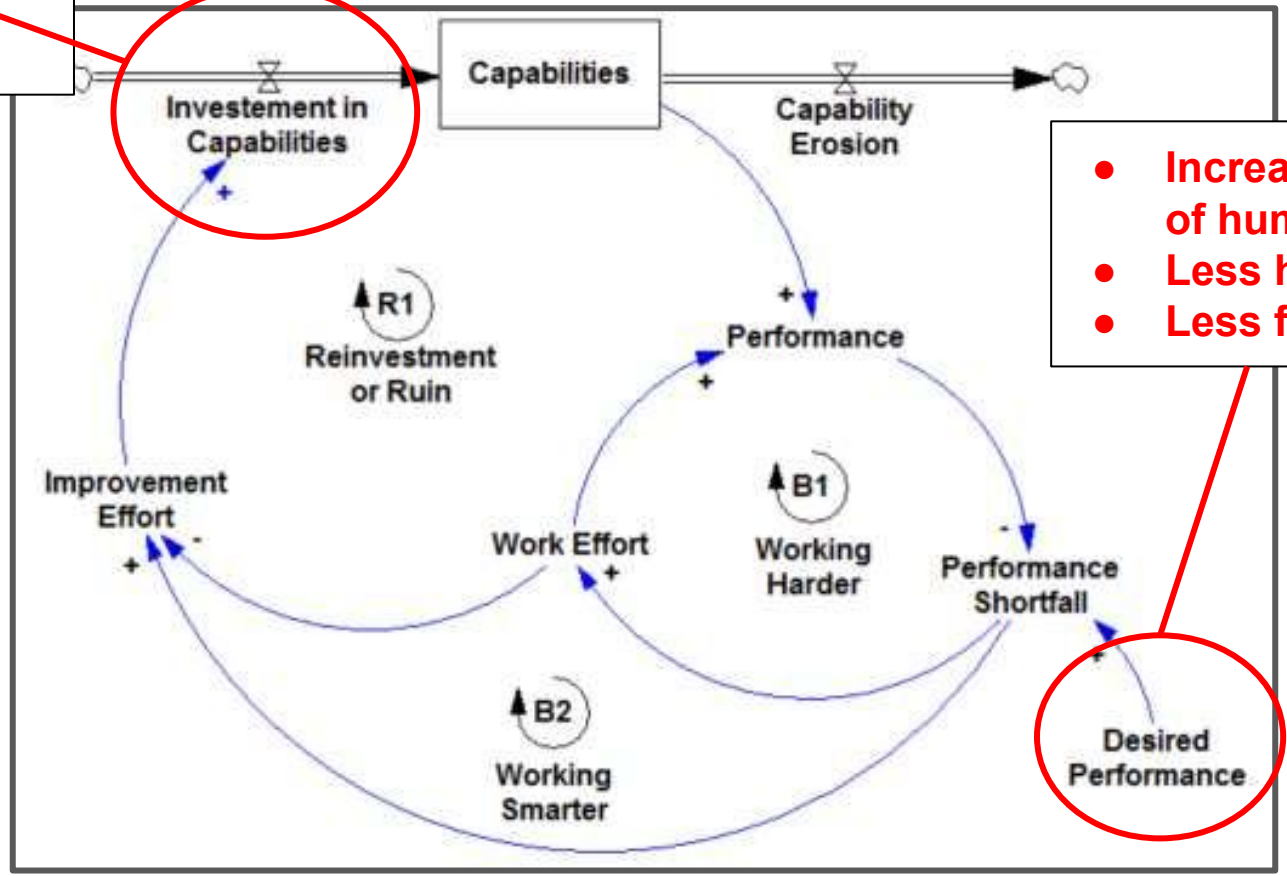
In order to combat the tendency to abstract away social context, this paper seeks to re-frame social context as a socio-cultural layer — which we will refer to as *societal context* — of the complex environment in which all technical systems and the social actors that create and are affected by them, exist and interact. Specifically, we introduce and leverage the multidisciplinary complex adaptive system (CAS) theory to develop a taxonomy model of societal context.

Next, we leverage the taxonomy model and fundamentals of product development processes to propose the concept of collaborative causal theory formation (CCTF), which we identify as a needed capability for incorporating societal context into the ML system design process in partnership with other (often excluded) stakeholders. We focus particularly on operationalizing CCTF through the use of system dynamics (SD) [34, 101], which is a transparent and rigorous visual and analytical tool for facilitating recursive engagement among diverse stakeholders. In practice, SD and bottom-up variants such as community based system dynamics (CBSD) are analogous to other efforts in the ML fairness community [7, 111] seeking to aid developers and researchers who desire working as partners with impacted stakeholders to develop greater perspective on the social and ethical dimensions of their research and products.

[1] shorturl.at/ouJO4
[2] shorturl.at/bgxCK

SD → Complex Problem Modeling

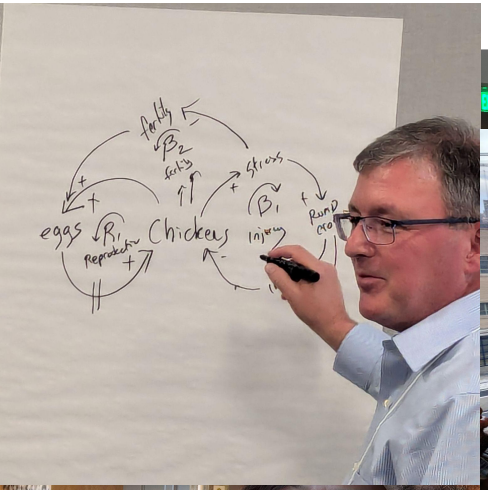
The AI Capability Trap

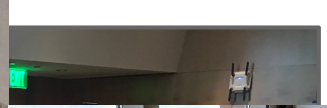


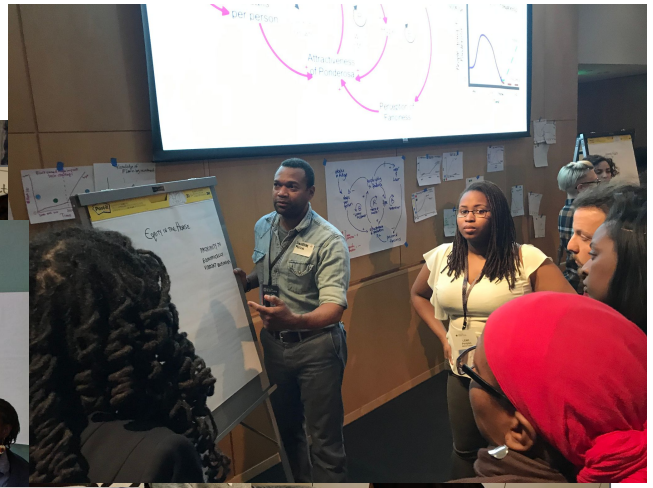
- **Increased visibility of humanity**
- **Less harmful bias**
- **Less fragile systems**

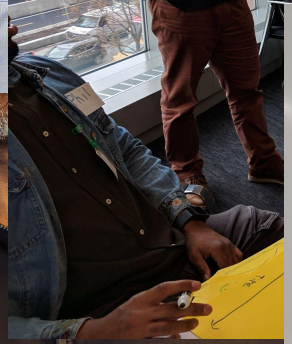
Base Diagram: Landry, E. and Sterman, J., 2017, July. The capability trap: prevalence in human systems. In 35th international conference of the system dynamics society (pp. 963-1010). System Dynamics Society.

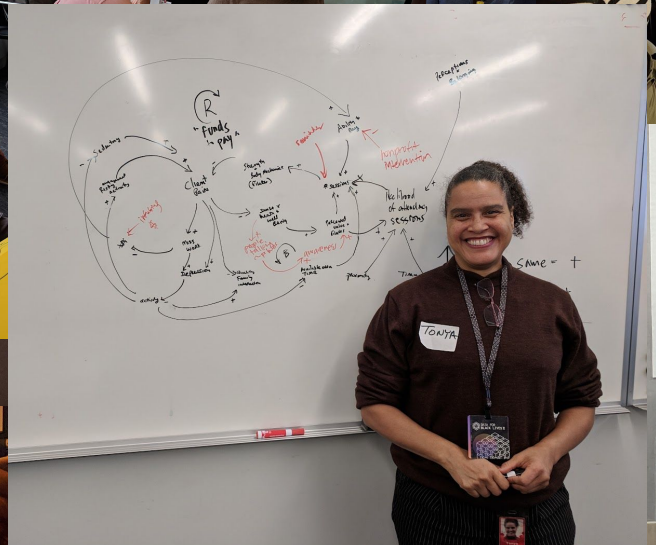
7 years of capability building effort …

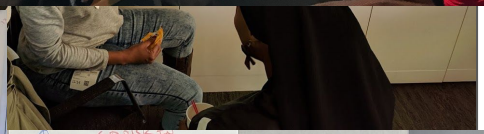


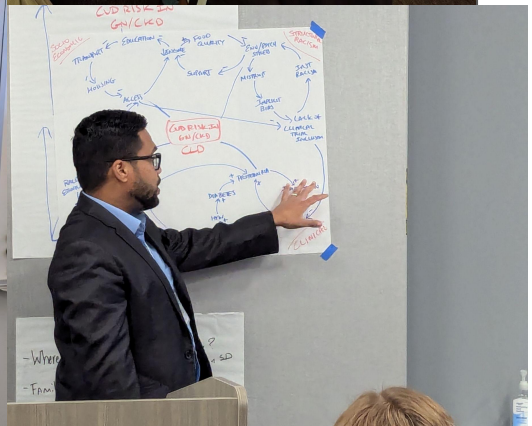












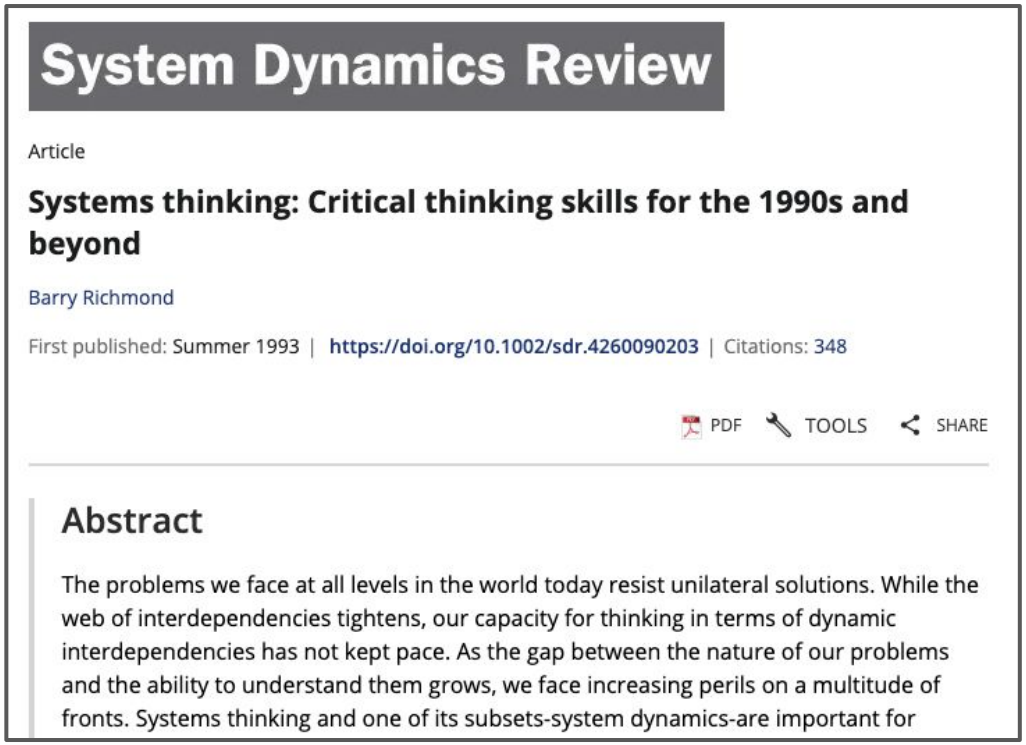
‘I really love the systems thinking approach. It has huge potential to transform product management. But, after leaving the workshop I discovered **it was way too hard to integrate system dynamics into my workflows or the workflows of my colleagues**.

The tools are **just too expensive and hard to use** and t**he amount of overhead it would take for my team to learn and integrate the concepts is just too high.**’

May 2019, very seasoned big tech product manager.

How AI Can Help SD

… if we’re careful.



System Dynamics Review

Article

Systems thinking: Critical thinking skills for the 1990s and beyond

Barry Richmond

First published: Summer 1993 | https://doi.org/10.1002/sdr.4260090203 | Citations: 348

PDF TOOLS SHARE

Abstract

The problems we face at all levels in the world today resist unilateral solutions. While the web of interdependencies tightens, our capacity for thinking in terms of dynamic interdependencies has not kept pace. As the gap between the nature of our problems and the ability to understand them grows, we face increasing perils on a multitude of fronts. Systems thinking and one of its subsets-system dynamics-are important for

"Unfortunately, we as System Dynamicists and Systems Thinkers have been **woefully inadequate in transferring our framework, skills and technologies to the population at large.**"

The billion dollar question

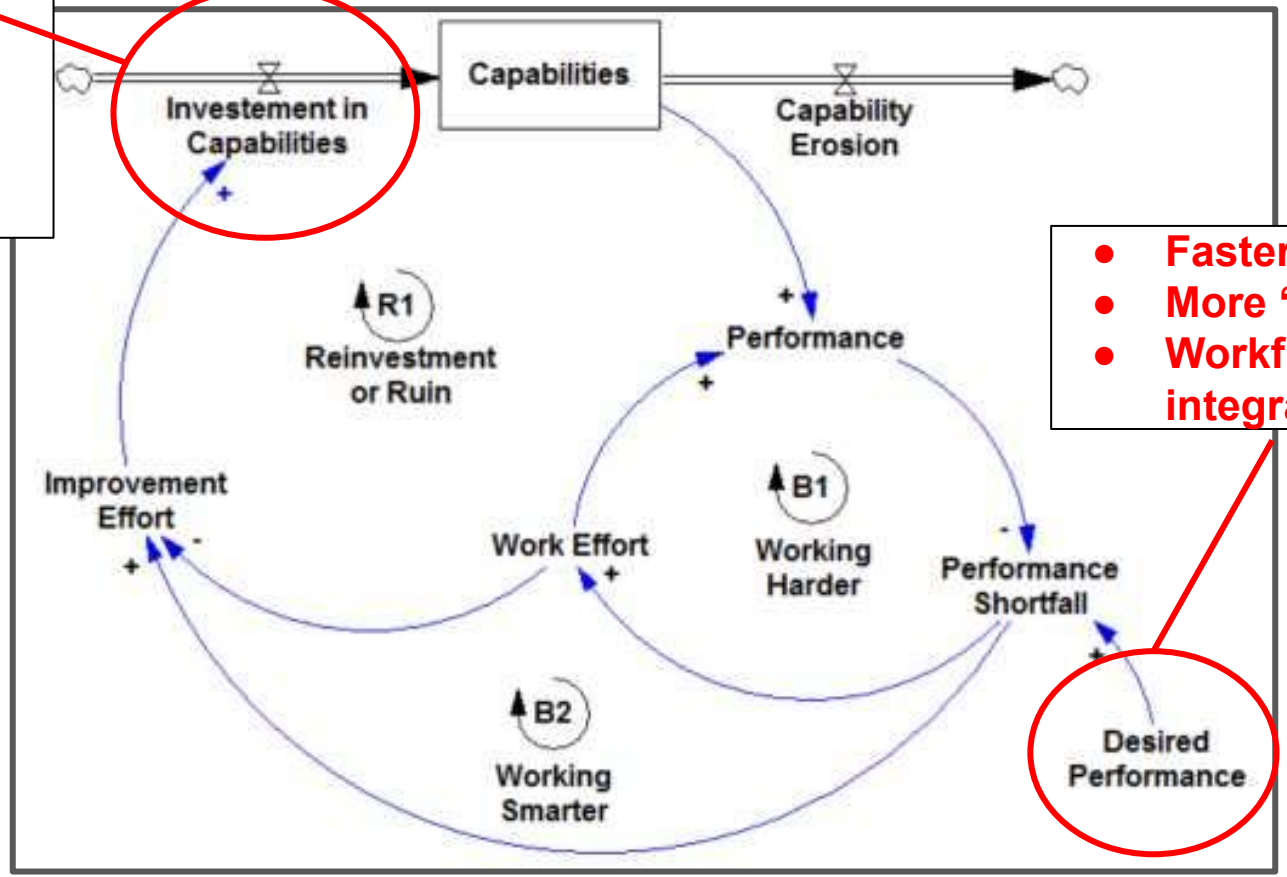
“How can the **framework, process and technologies of systems thinking be transferred to the rest of the world in an amount of time that is considerably less** than what it currently takes to get a Master's or PhD degree in our field?”

Barry Richmond, Systems thinking: Critical Thinking skills for the 1990s and beyond, SD Review, 1993

The SD Capability Trap

GenAI →

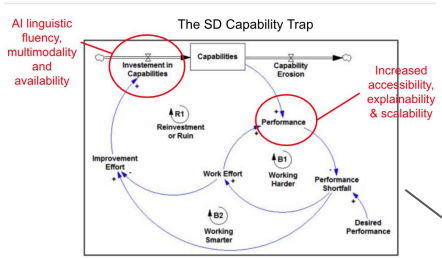
- **natural language**
- **multimodal**
- **scale**



- **Faster learning**
- **More “users”**
- **Workflow integration**

Base Diagram: Landry, E. and Sterman, J., 2017, July. The capability trap: prevalence in human systems. In 35th international conference of the system dynamics society (pp. 963-1010). System Dynamics Society.

Dueling capability traps …

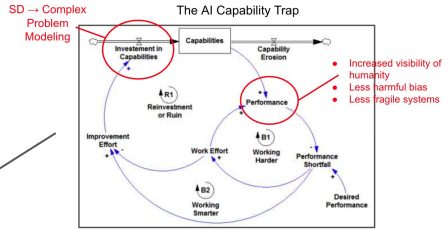


Can improve …

safety, robustness and efficacy of…

SD

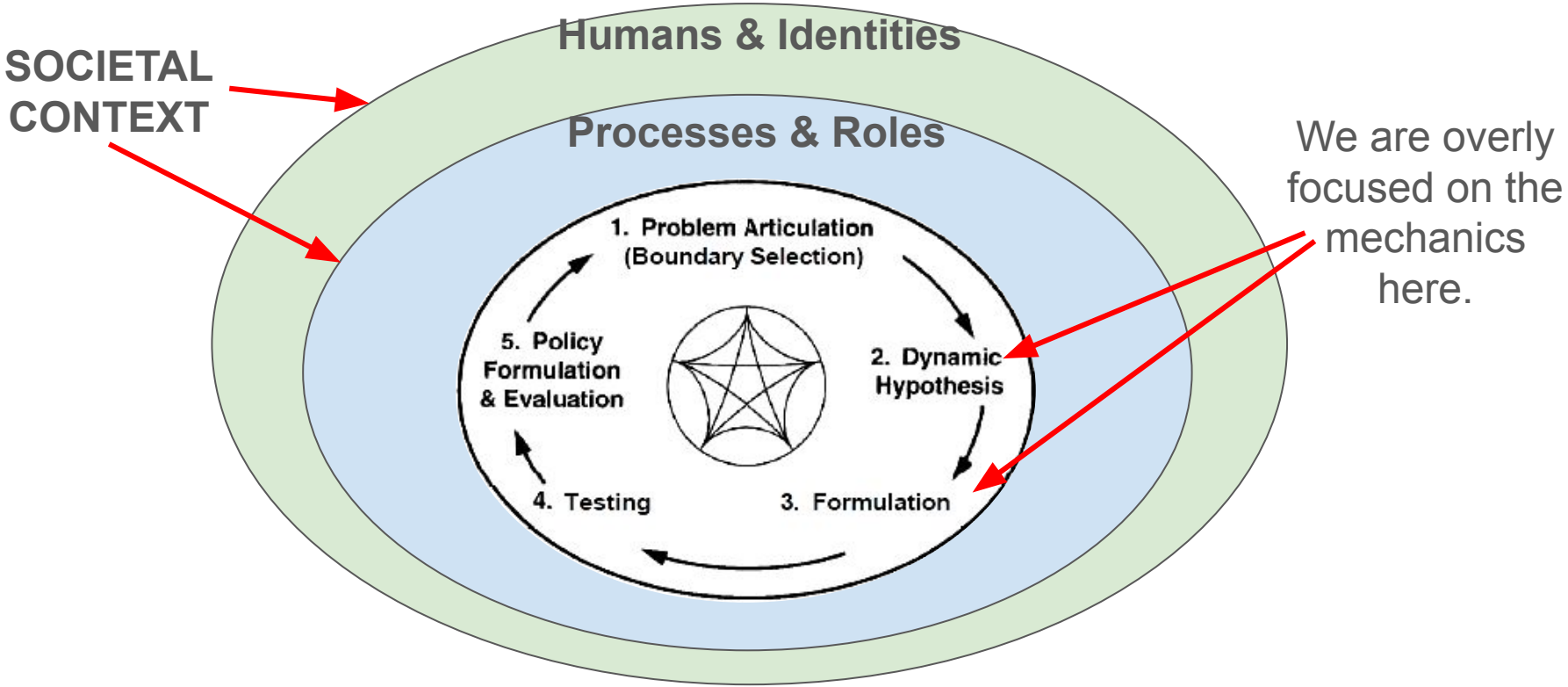
AI



… demand for, scalability, and accessibility of …

Can improve …

GenAI to escape the SD capability trap?: **if we're careful**



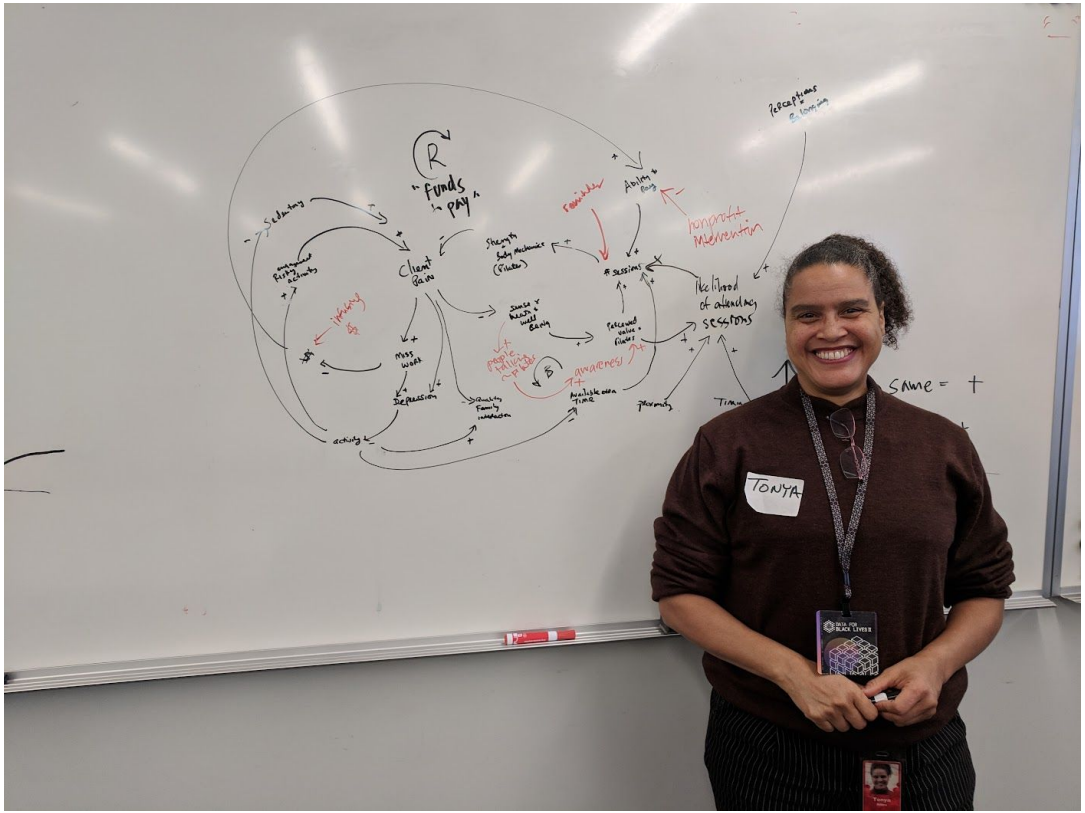
Base Diagram from Steps of SD modeling (Sterman, 2000)

GenAI to escape SD capability trap: **requires grounding**

- **Goal:** significantly increase the number and diversity of humans in the general public that leverage SD on a daily basis to get their core work done

- **Principles**
  - Prioritize situational, context-specific, learner-driven teaching and assistance
  - Help specific human roles and identities get specific jobs done in specific contexts
  - Focus on human capabilities first. Model capabilities are to augment and assist humans **- they are the means, not the end.**
  - Remember that group model building is the backbone of the practice

Staying grounded



Tonya Marie Amos, 2019, MIT, Boston, MA

Questions?

bit.ly/sdspolls