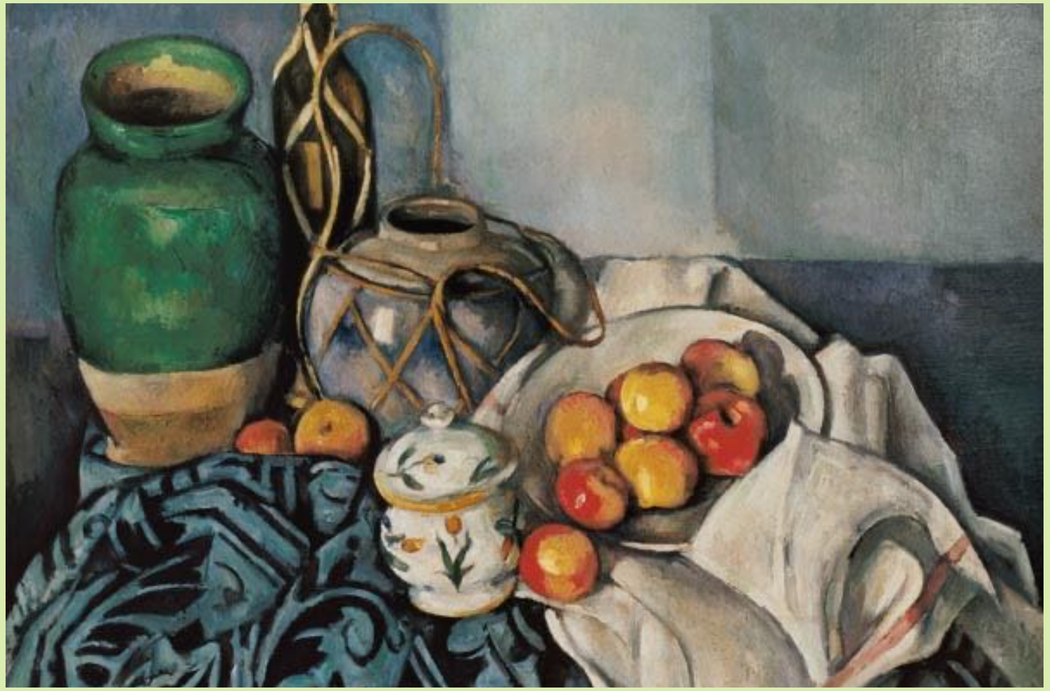
The Ontology Machine

that needs help



Johannes
Himmelreich
03/27/2026


“Ground Truth”

Two senses of “ground truth”

Both claim to be *veridical*

Platonic

Sortal, structured, propositional

Edenic

Simple, primitive, phenomenal

Ground Truth

Platonic

Example of a platonic dataset

CHILDID	X_CHSEX_R	X_RACETH_R	X1MSCALK5	X1RSCALK5	X12SESL
100234	1 (Male)	3 (Hispanic)	28.47	34.12	-0.68
100891	2 (Female)	1 (White, non-Hisp.)	42.91	51.03	0.84
101407	2 (Female)	2 (Black, non-Hisp.)	31.55	38.29	-0.31
102063	1 (Male)	4 (Asian, non-Hisp.)	46.18	44.76	1.12

Ground Truth

Example of an edenic dataset

Edenic


```
{"text": "The mitochondria are membrane-bound organelles found in the cytoplasm of eukaryotic cells. They generate most of the cell's supply of adenosine triphosphate (ATP), which is used as a source of chemical energy. Mitochondria have their own DNA, which is circular and encodes 37 genes.", "meta": {"pile_set_name": "Wikipedia (en)"}}

{"text": "Abstract: We present a novel approach to low-rank matrix factorization that achieves sublinear time complexity under mild incoherence assumptions. Our method leverages randomized sketching to approximate the top-k singular vectors without forming the full matrix. Experiments on synthetic and real-world datasets demonstrate a 3x speedup over existing methods with negligible loss in reconstruction error.", "meta": {"pile_set_name": "ArXiv"}}

{"text": "def quicksort(arr):\n    if len(arr) <= 1:\n        return arr\n    pivot = arr[len(arr) // 2]\n    left = [x for x in arr if x < pivot]\n    middle = [x for x in arr if x == pivot]\n    right = [x for x in arr if x > pivot]\n    return quicksort(left) + middle + quicksort(right)", "meta": {"pile_set_name": "Github"}}

{"text": "UNITED STATES DISTRICT COURT\nSOUTHERN DISTRICT OF NEW YORK\n\nJANE DOE, Plaintiff,\nv.\nACME CORP., Defendant.\n\nNo. 19-cv-04472\n\nORDER\n\nDefendant's motion to dismiss for failure to state a claim under Fed. R. Civ. P. 12(b)(6) is DENIED. The complaint adequately alleges that Defendant's conduct constituted a breach of the implied warranty of merchantability under N.Y. U.C.C. § 2-314.", "meta": {"pile_set_name": "FreeLaw"}}

{"text": "Hey everyone, I've been running a 5e campaign for about 8 months now and my players just hit level 12. The problem is the wizard has become absurdly powerful compared to the rest of the party. Wall of Force alone trivializes most encounters. Any DMs have advice on how to handle this without nerfing the player directly?", "meta": {"pile_set_name": "OpenWebText2"}}
```

Ground Truth

Edenic

Example of an
edenic dataset

Ground Truth

Example of an eden-*ish* dataset

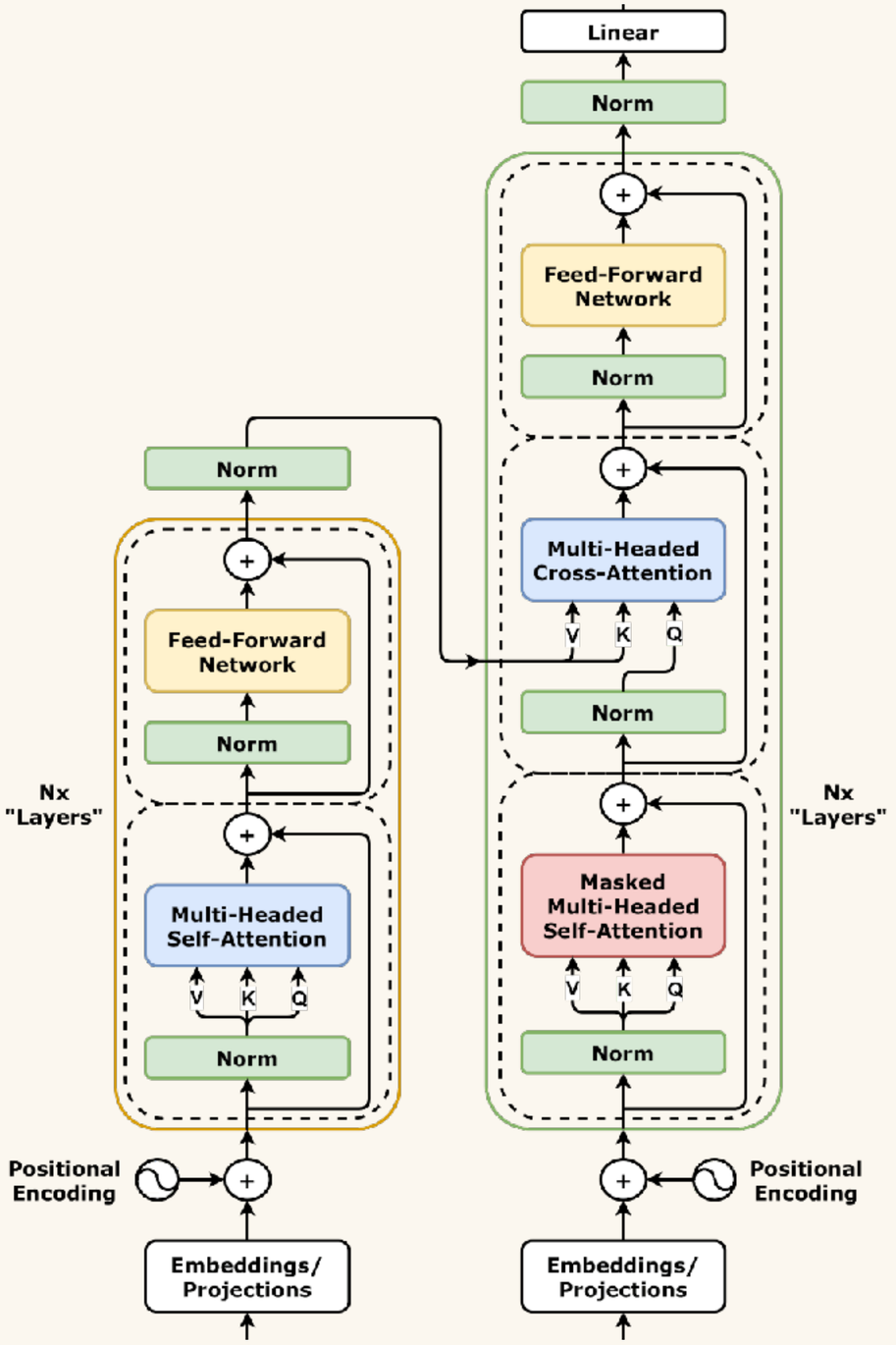
Eden-*ish*

Modal AI has been trending towards the edenic myth

Neural networks as an “ontology machine”

Unstructured data

Learns features



AI needs ontological help

Edenic AI has an “ontological bottleneck”

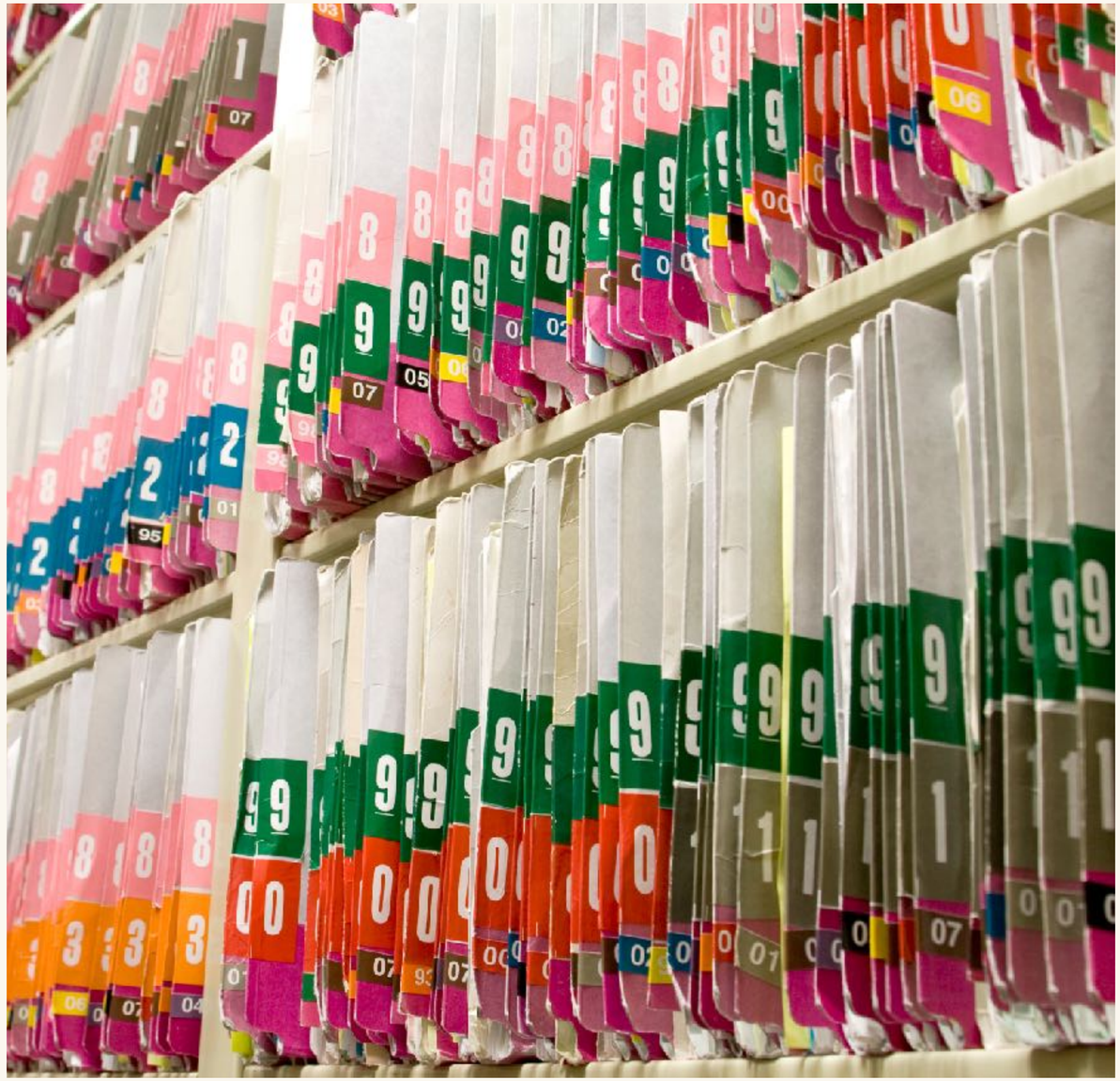
Health

Construct

Operationalized via *care expenditures*

Source of bias: Label bias

- **Construct bias (HEALTH?)**
- **Measurement bias**



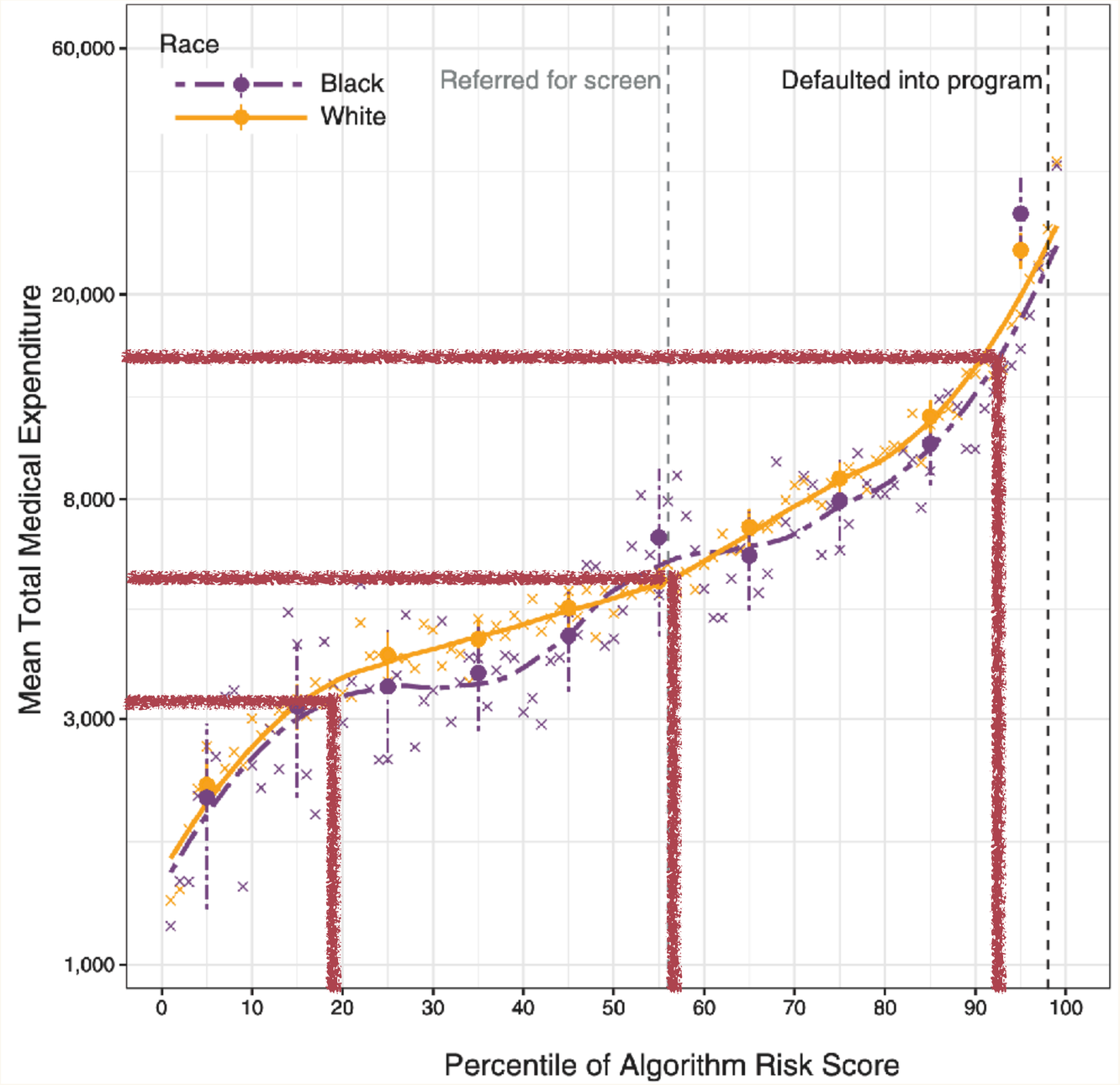
Health

Construct

Operationalized via *care expenditures*

Source of bias: Label bias

- **Construct bias (HEALTH?)**
- **Measurement bias**



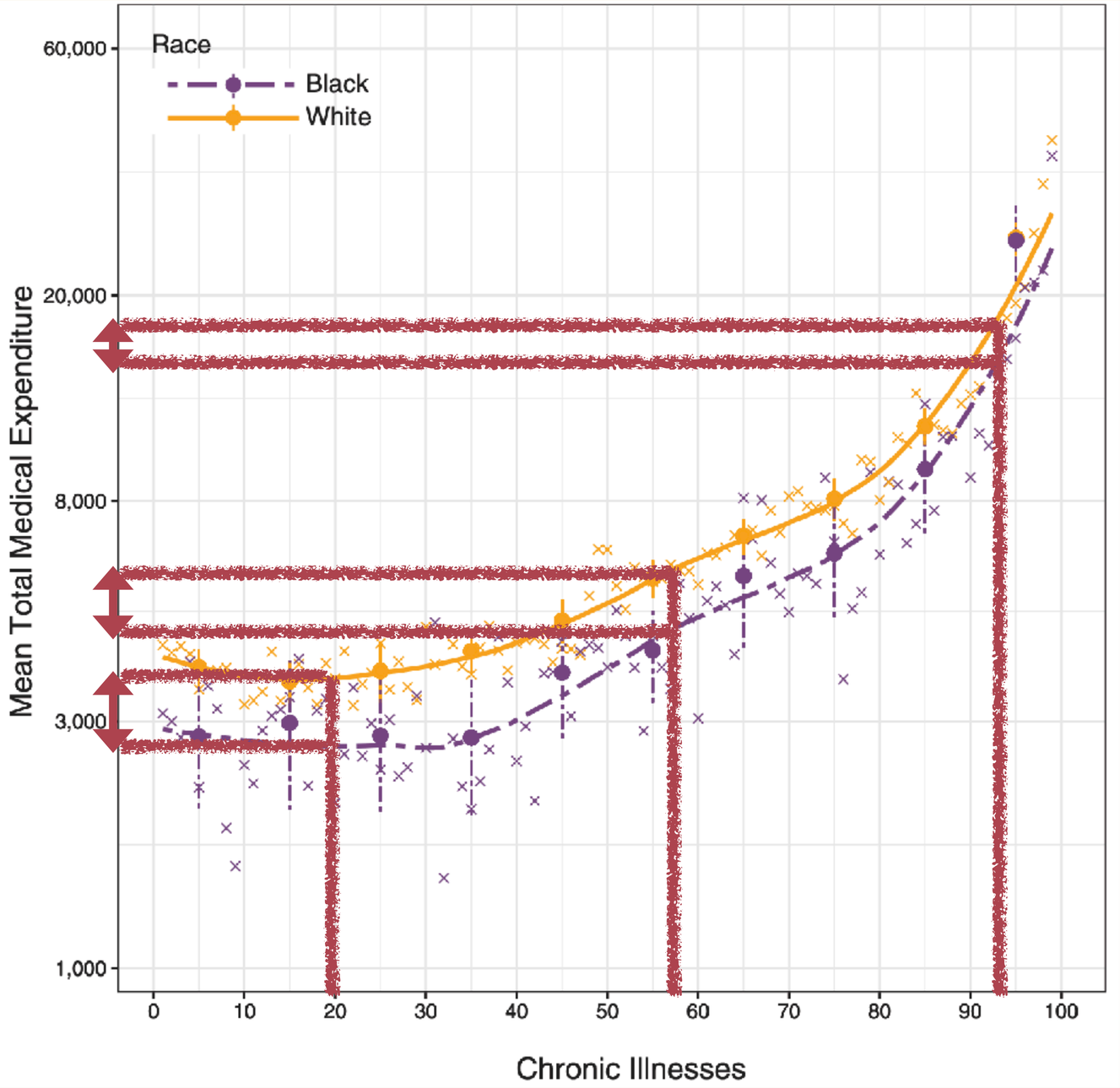
Health

Construct

Operationalized via *care expenditures*

Source of bias: Label bias

- **Construct bias (HEALTH?)**
- **Measurement bias**



Current AI has an “ontology bottleneck”

Explicating visual features is hard

Neural networks learn these

Explicating social constructs is easy

Doing so *well* is hard (what is that even?)

AI rarely makes use of social constructs

They are not in the data or benchmark

Example 1

Automatic transformation of narrative orientation

What the transformer doesn’t know

Subtextual properties

Narrative cultural orientation

“But it will not rise for the feet of a few. *Only when all hands dig—young, old, weak, strong—will the spring awaken.*”

“But it will not rise for the feet of a few. *Only when one determined soul digs will the spring awaken.*”

Example 2

Construct here is *implicit*: not explicated or operationalized

What the transformer doesn’t know

Subtextual properties

Ethical perspective

Care ethics

Act consequentialism

Several problems

Construct bias

Label and measurement bias

Conceptual diversity

Metaphysics—Epistemology trade-off

Conceptual dynamics

Ideas change extension and intension

Adapting concepts

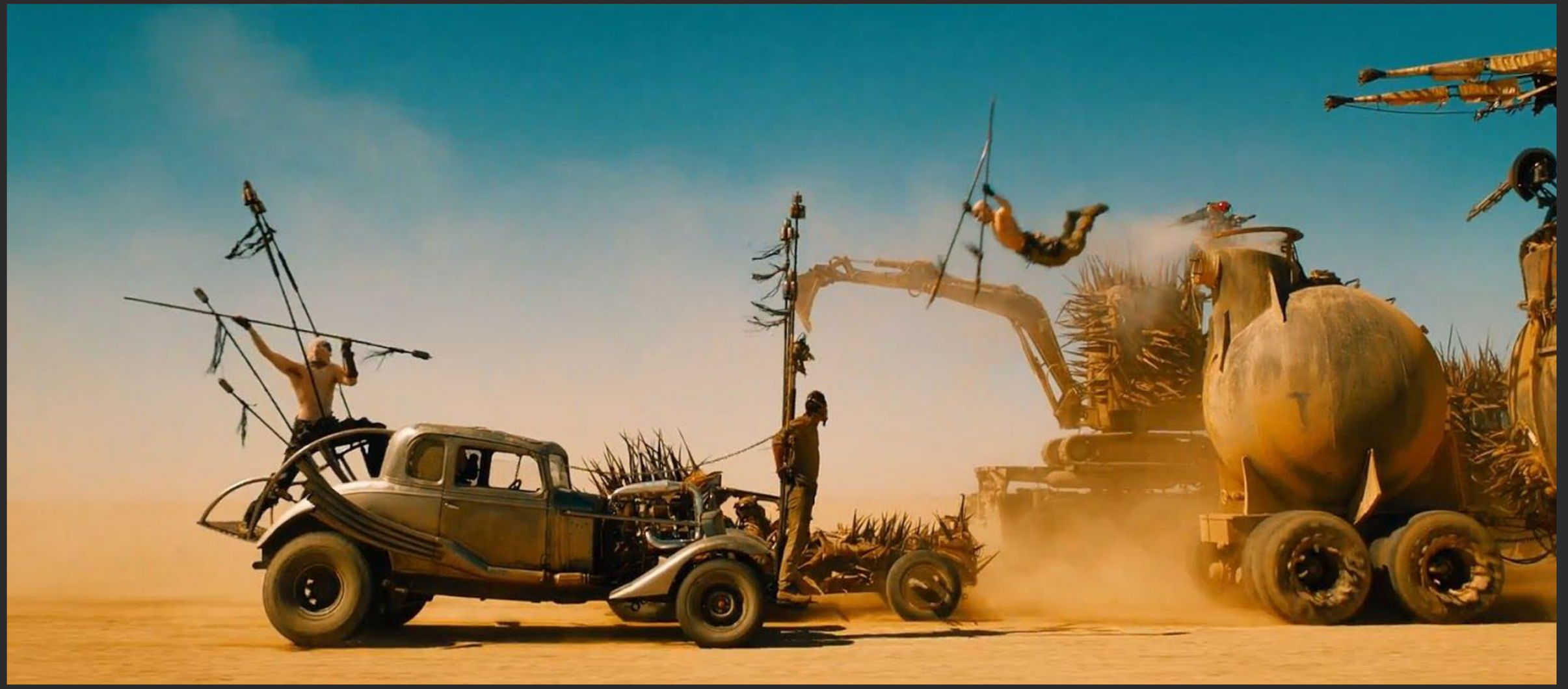
Over time

Type	Mechanism	Examples
Creation	Name what was nameless	Sexual harassment, genocide, epistemic injustice, global domestic politics, popup geopolitics
Revision	Redefine to expose	Woman, race, misogyny, rape, disability
Extension	Broaden who counts	Person, citizen, marriage
Reclamation	Invert valence of stigma	Queer, Black

Adapting concepts

To AI

Type	Mechanism	Examples
Creation	Name what was nameless	Vibe coding, …
Revision	Redefine to expose	orange–Teal contrast, average
Extension	Broaden who counts	agent
Reclamation	Invert valence of stigma	?



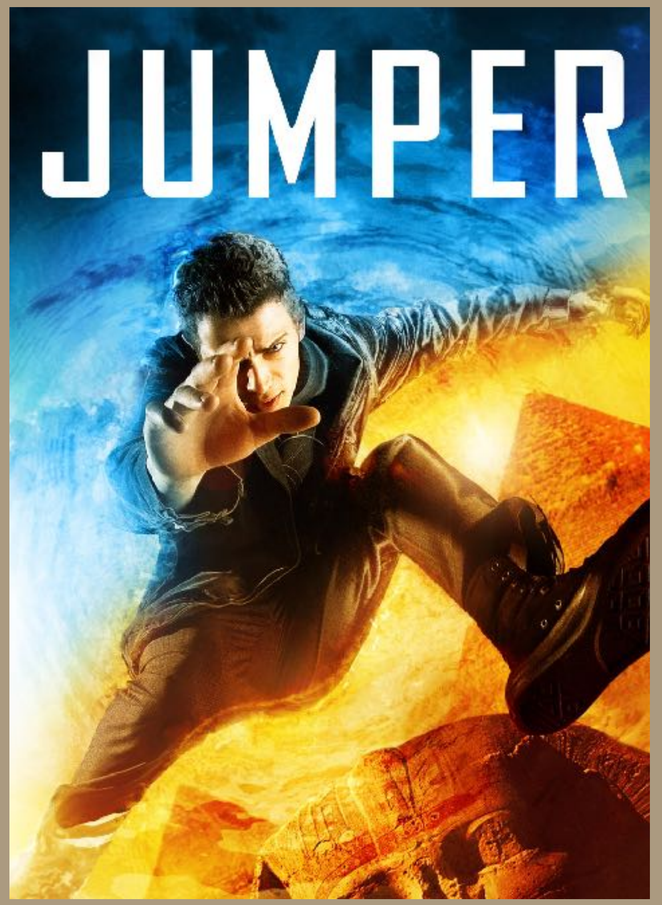
Orange–Teal

We have a name for this, everyone can see it, but post-production editors have no shame

Mad Max (2015)

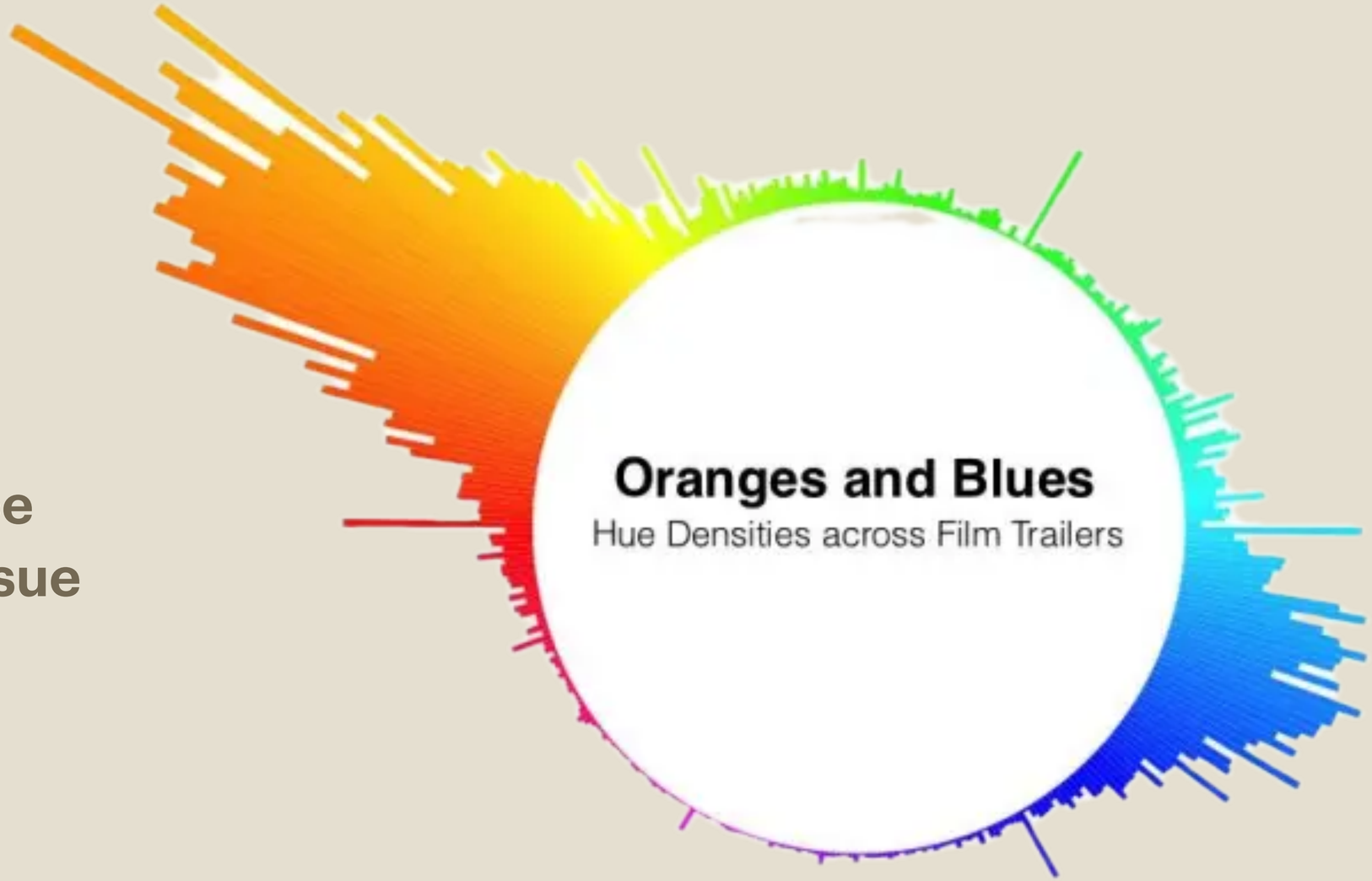
G.I. Joe (2009)

Jumper (2008)



It’s real

Orange–teal contrast: candidate concept to be revised to expose an issue



Concluding remarks

Being human in the age of AI
Being AI in the age of humans

Summary

Or something close enough

Current AI is an ontology machine

That was the point of neural networks

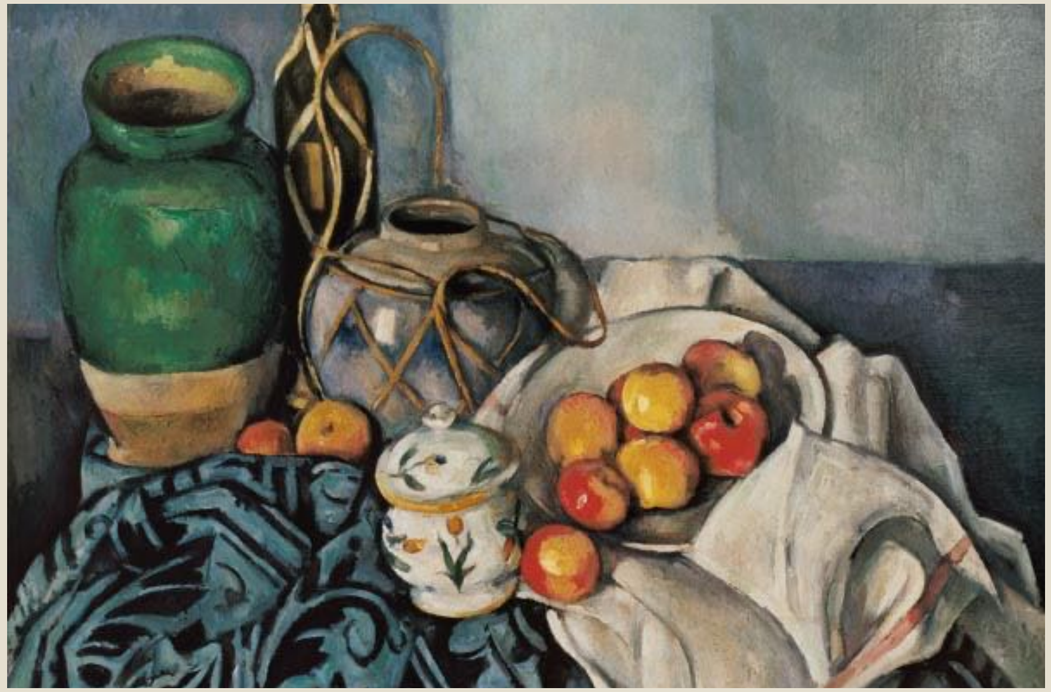
So are we (individuals, groups, society)

Ontology = understanding, sense-making

***We*—AI are not ontologically linked up**

A technical problem but mostly not just a technical problem

Not a problem even, it’s human



Johannes Himmelreich
Associate Professor
Syracuse University
jrhimmel@syr.edu