B. Scot Rousse
Topos Oxford Seminar
June 19, 2025

Based on work in progress with David Spivak

An Exercise

➢ "Values" is an abstraction that is rooted in a deeper and more pervasive phenomenon–care.

➢ Values are abstracted extractions from caring that can detach us from our caring.

➢ We need to "take care" when we appeal to values.

➢ We always need to ask "*Who cares* about these values?"

Caring: a pre-cognitive attunement to significance that resists complete formalization while enabling responsive involvement and judgment.

Caring:

➢ Not just a feeling of affection.

➢ The activity of *tending and attending to what matters*.

➢ An attuning and cultivating that tends to bring things out at their best.
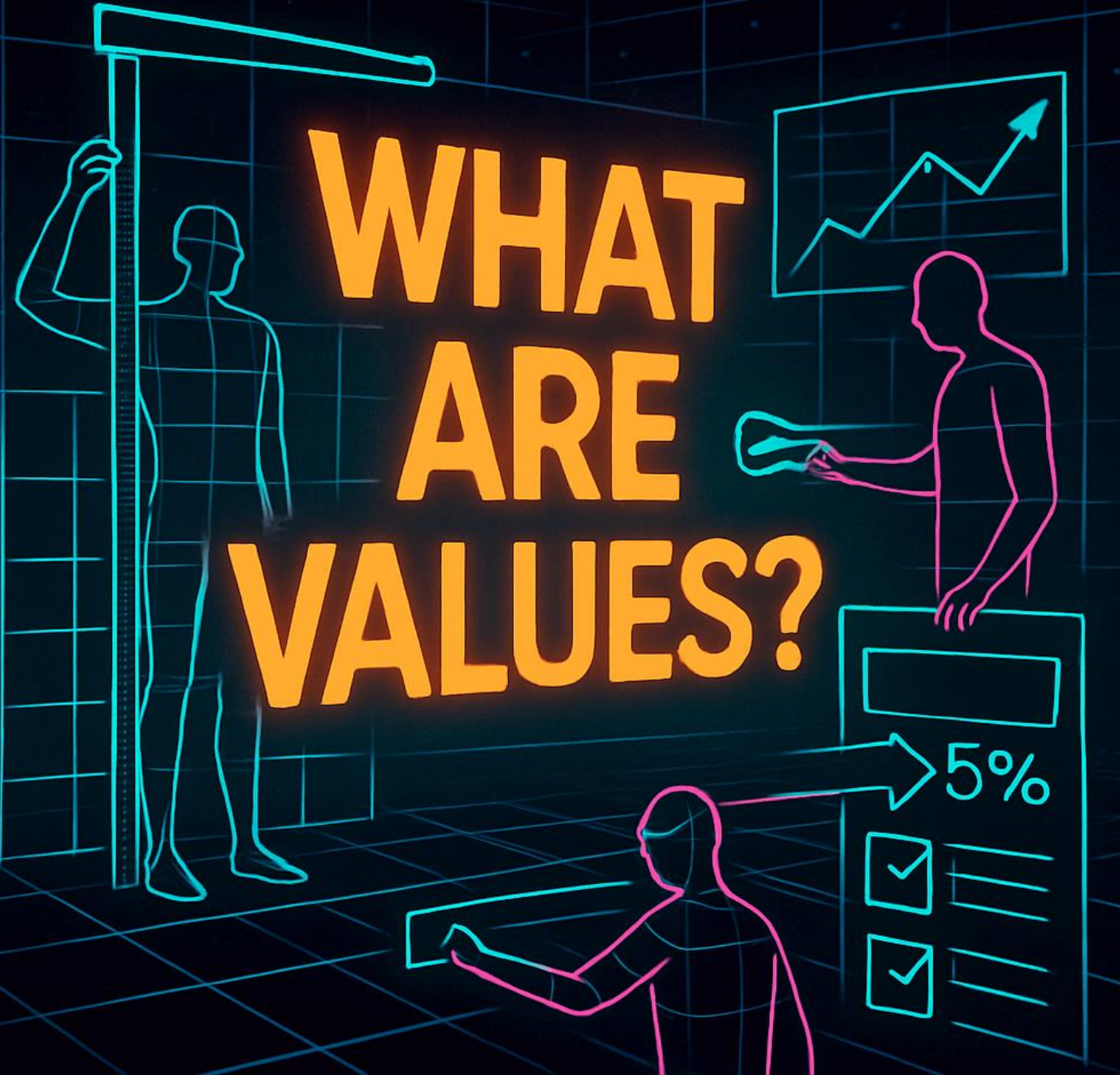
➢ "Caring tends to actualize potential."

"Values" attempt to grasp and translate the importance of what we care about into

- a precisely legible
- quantifiable

→ An inert entry – a *value!* – on a spreadsheet

→ "The Legibility Pretense"

The fact that values can be translated to the quantified value in a spreadsheet is certainly part of their usefulness.

Statements of "our values" help us make plans, calculate between alternatives, and control for and evaluate specific outcomes.

What if the notion of "values" is an inadequate abstraction to capture the importance of what we care about?

→ The "specification problem" is not just a technical shortfall: it reveals a fundamental flaw in using "values" to translate care.
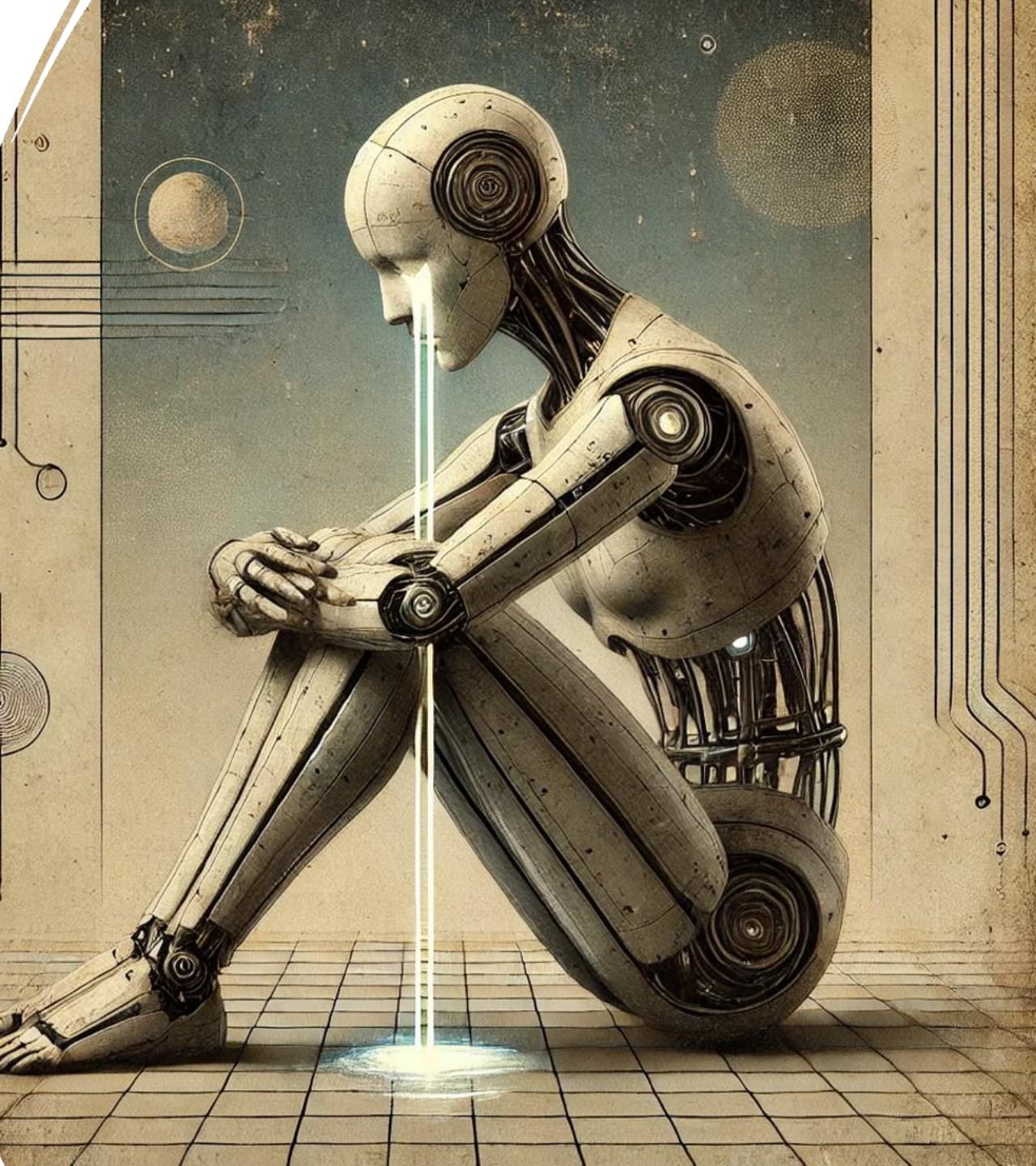
Values are abstractions (or extractions) from caring that we often forget are abstractions, and that we then relate to as fixed realities and data points.

➢ This can detach us from the care that the values were meant to capture.

➢ The provisional nature of values is one of the most important aspects to keep clear when letting values get to work in a discussion.
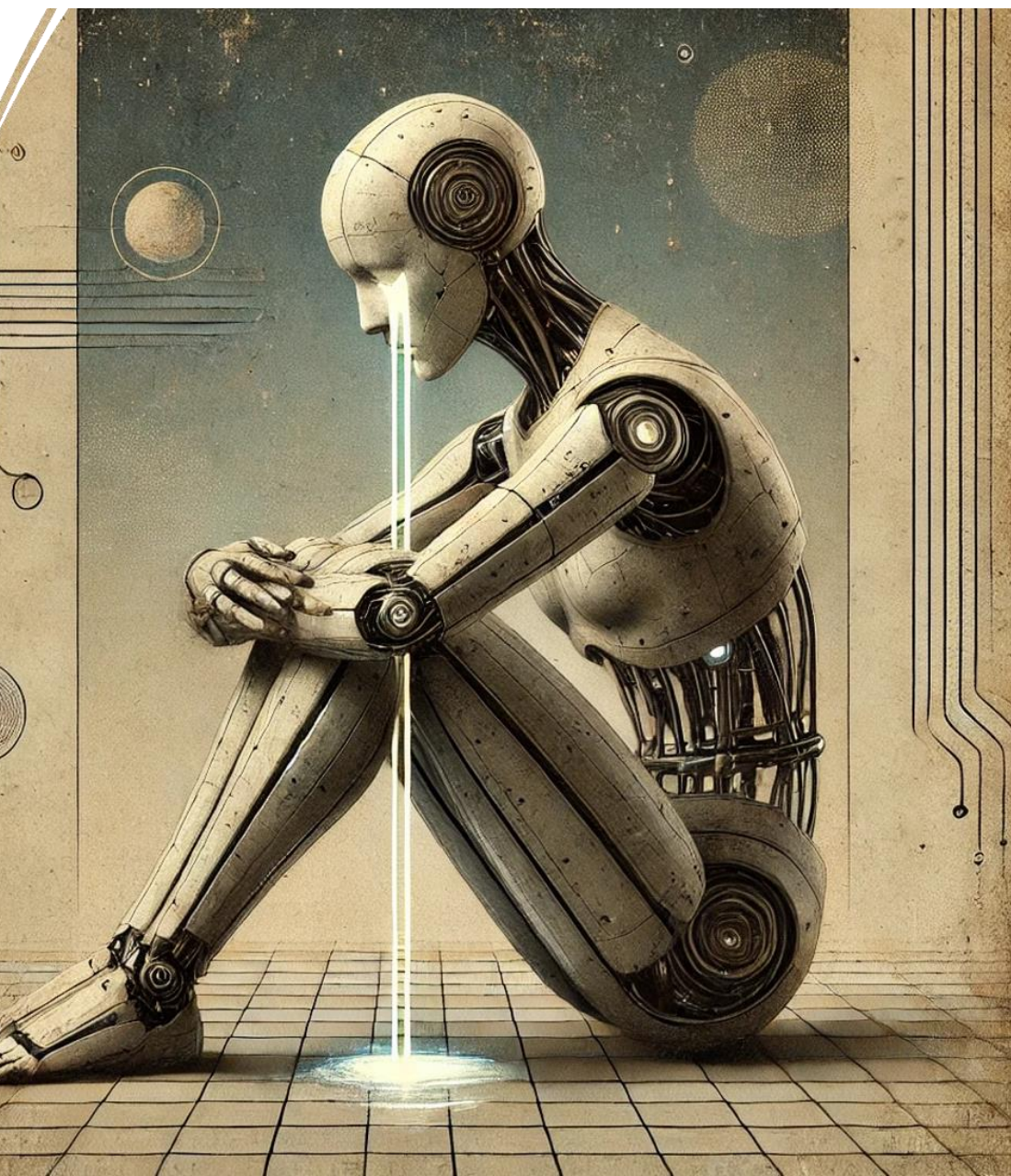
Values are abstracted extractions from caring that can detach us from our caring

➢ The preference boxes on a dating app that sort us into stereotypical types or "profiles," rendering what we care about into data for the algorithm.

➢ A corporation's values that become hypostasized, creating harm, e.g., Amazon's amazing value of customer satisfaction giving rise to dangerous or humiliating worker conditions.

The Historical Emergence of Values:

Two characteristics:
➢ Objectifying
➢ Subjectifying

The emergence of our contemporary notion of "value" in 19th century thought

[I] Values as Objectifying

Naturwissenschaften (natural sciences)

Geisteswissenschaften (human sciences)

Weber:

"Value – Relevance" vs "Value-Neutrality"

Max Weber (1918)

The emergence of "value" in 19th century thought

II. Values as "Subjective"

The notion of "*values* [brings] with it the assumptions that all moral ideas are subjective and relative…

…That they are mere customs and conventions, that they have a purely instrumental, utilitarian purpose, and that they are peculiar to specific individuals and societies.

Gertrude Himmelfarb, *The De-moralization of Society*, 1994, p.11.

The emergence of "value" in 19th century thought

[II] Values as "Subjective"

Today we see values as items on a check-lists of preferences that individuals (or organizations) have or possess in some way…

We "have" values, rank them on a table, where them on a shirt, edit and delete them as we see fit.
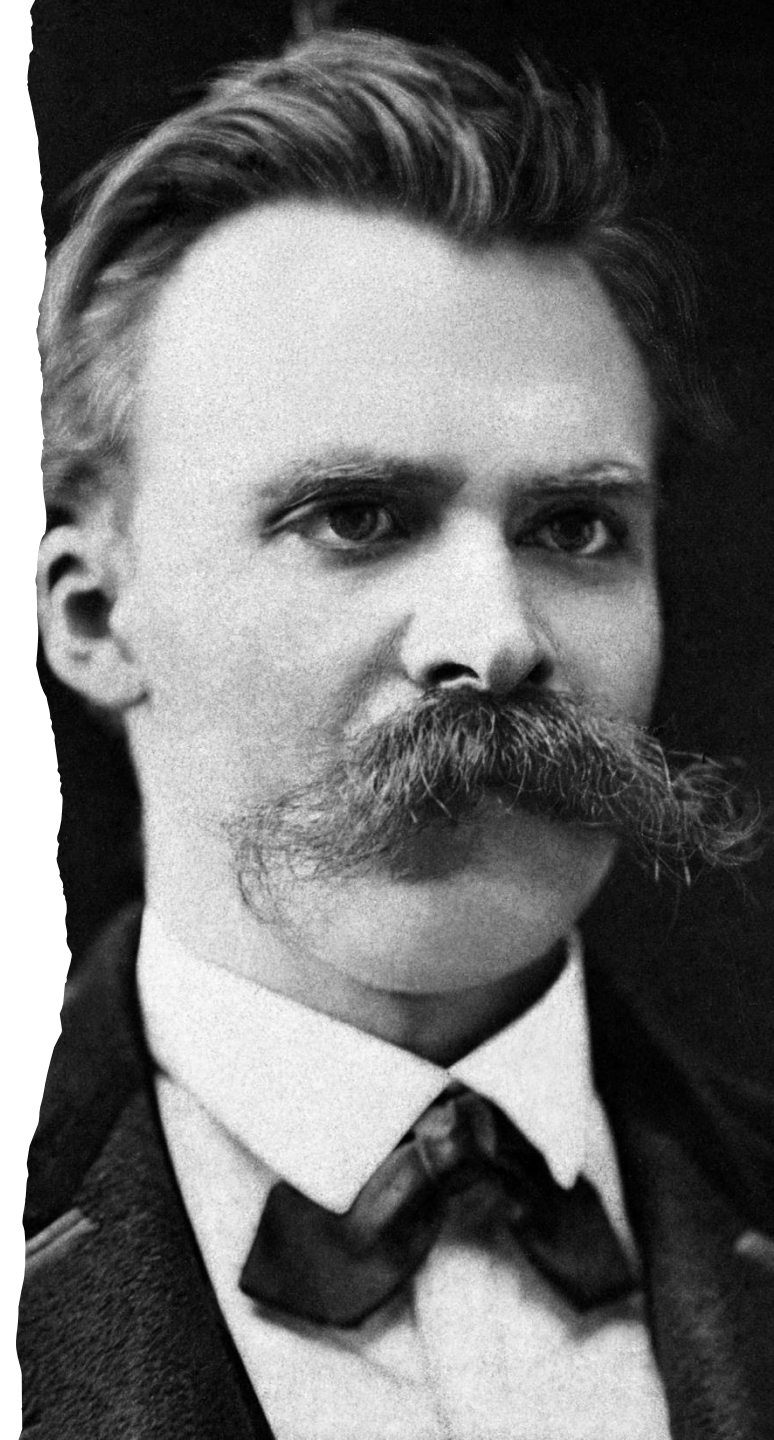
Values: something we *have*
Caring: something we *do*

The concept of "value," in both its objectivizing and subjectivizing dimensions, enacts an "outside-in," or "sideways-on" observer's perspective on human life.

→ Our concept of "value" today retains these connotations which influence our thinking and conducting of human life accordingly.

Living in a technological age:

→ Compulsive efficiency and optimization

→ We relate to life (and the world) as a series of problems to solve.

→ An orientation to "values" is part of this picture.

Martin Heidegger, "The Question Concerning Technology" (1954)

A focus on values easily lends itself to a narrow, utility-maximization perspective that attracts our attention, and gives us powers of prediction and control…

…but risks turning us away from what really matters.

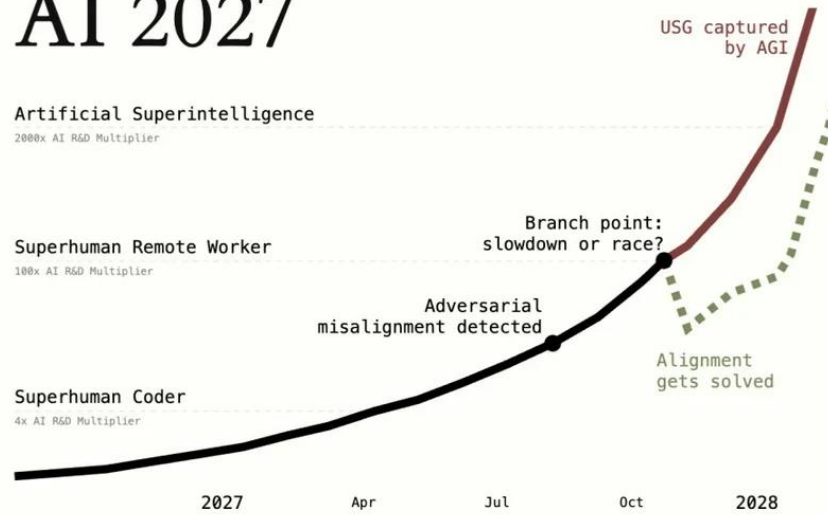*You cannot properly tend to what matters while trying to optimize and control it.*

Cf, factory Farming
industrialized agriculture
transactional friendships

"[W]e do have an ASI over here, answering all of our questions, doing whatever we want. You would think that people could just ask, 'Hey, Superintelligent AI, where does this lead? Or, 'What happens?' Or, 'How is this going to affect human flourishing?'

And then it says, 'Oh yeah, this is terrible for human flourishing, you should do this other thing instead.' ...

...If we know with certainty, because the AI tells us, that this is just a stupid way way to do everything, **is less efficient, makes people miserable,** is that enough to get the political will to do the UBI or not?"

- Scott Alexander to Dwarkesh Patel (2025)

THE LAST QUESTION

- Isaac Asimov, "The Last Question" (1956)

"Ask the Microvac," wailed Jerrodette I. "Ask him how to turn the stars on again."

"Go ahead," said Jerrodine. "It will quiet them down." (Jerrodette II was beginning to cry, also.)

Jerrodd shrugged. "Now, now, honeys. I'll ask Microvac. Don't worry, he'll tell us." He asked the Microvac, adding quickly, "Print the answer."

Jerrodd cupped the strip or thin cellufilm and said cheerfully, "See now, the Microvac says it will take care of everything when the time comes so don't worry."

Jerrodine said, "And now, children, it's time for bed. We'll be in our new home soon."

Jerrodd read the words on the cellufilm again before destroying it: INSUFICIENT DATA FOR MEANINGFUL ANSWER.
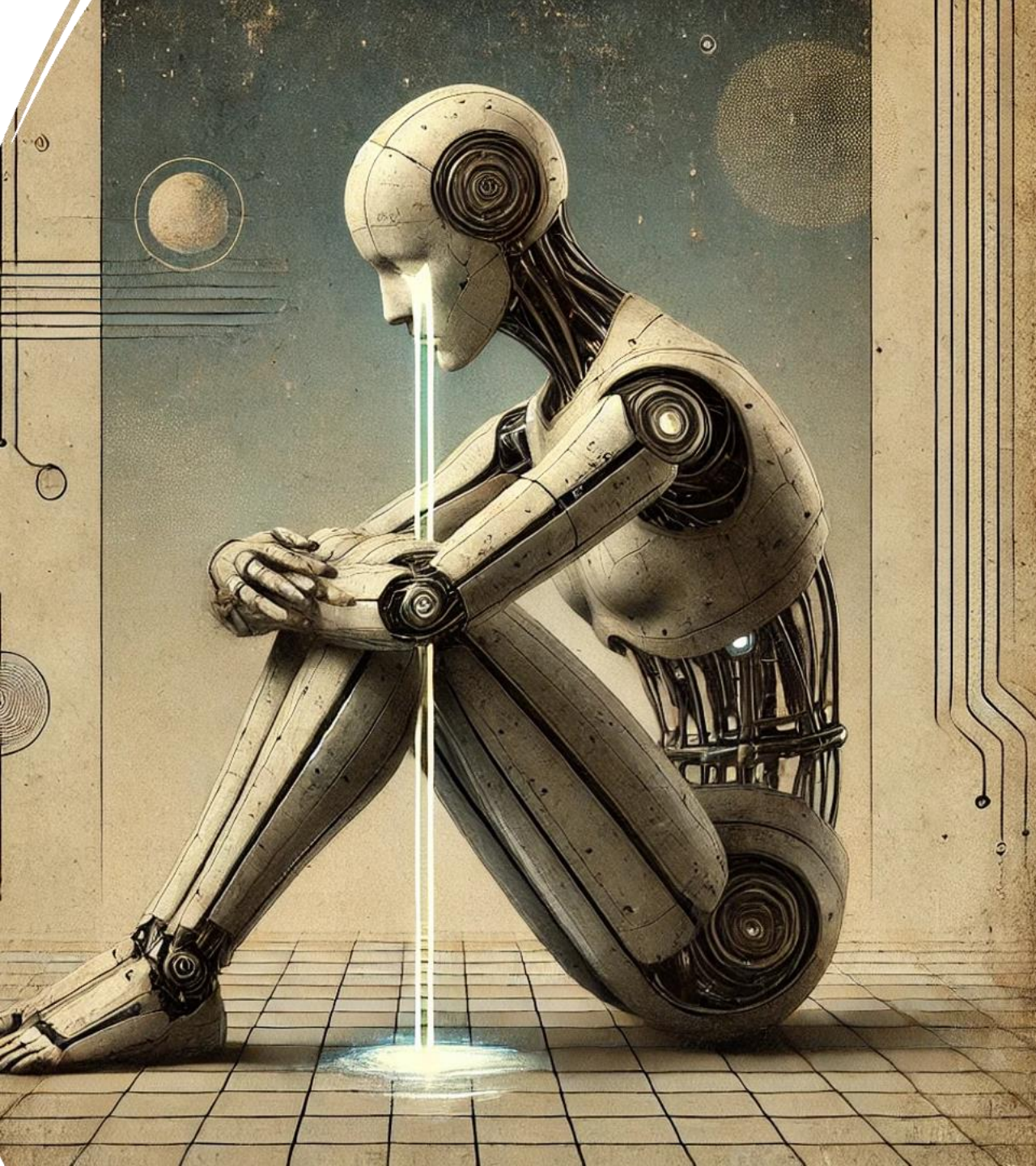
Values: Objectified representations of care that abstract from world-embedded responsiveness into measurable, comparable properties.

➤ **Valuing:** stepping back and out to measure and state what matters

Care: Pre-cognitive attunement to significance that resists complete formalization while enabling responsive involvement and judgment.

➤ **Caring**: being drawn to tend (and attend) to what matters

1. Distinguishes between what importance & triviality

2. Connected to vulnerability and mortality

3. Possibly exceeds our reflective awareness

4. Possibly exceeds reason & morality

5. Shapes our will (Luther's "Here I stand!"

6. Embedded in socially shared forms of understanding

7. Embodied in manners & dispositions

Caring provides our orientation in the world.

What we care about solicits our attention and action, involving us in the world with other.

When we care, we are drawn to tend and attend, to nurture, cultivate, bring a thing out at its best.

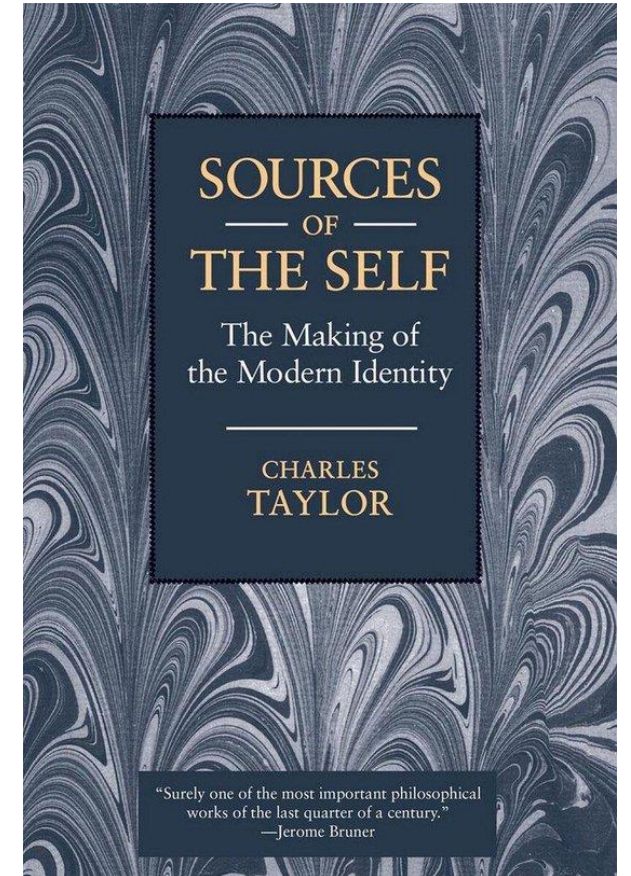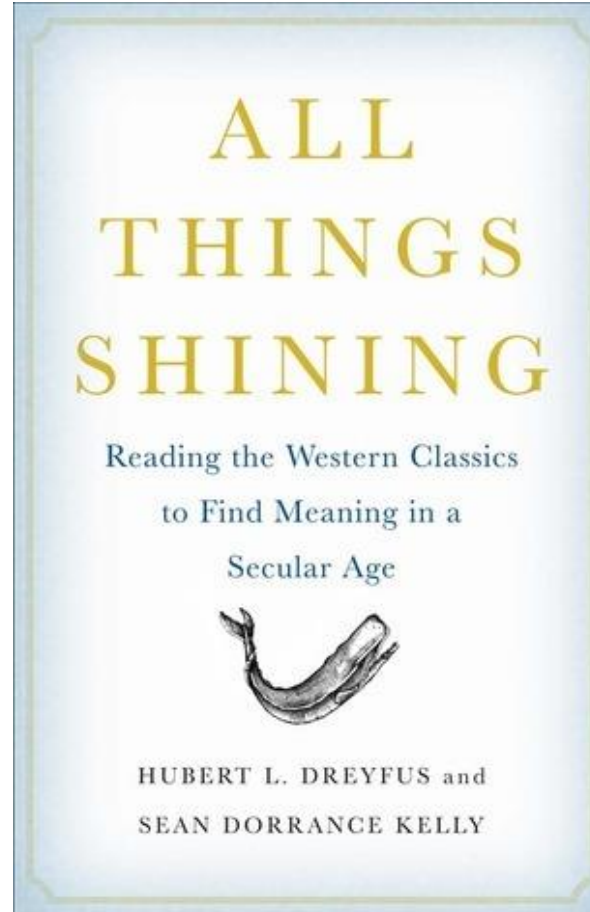… Which can sometimes pit us against reason, utility, common sense, and morality.

A potted history of care, well-being, and autonomy in the West

- Being appropriately cunning and receptive to the influence of the various gods (Homer)
- Justice as harmony in the soul of the individual & polis, lining up with cosmic order embedded in the forms (Plato)
- Eudaimonia in Aristotle
- Dante's alignment with the love of god structuring the cosmos

"Here powers failed my high imagination: But by now my desire and will were turned, Like a balanced wheel rotated evenly By the Love that moves the sun and the other stars."

ALL THINGS SHINING

Reading the Western Classics to Find Meaning in a Secular Age

HUBERT L. DREYFUS and SEAN DORRANCE KELLY

SOURCES OF THE SELF

The Making of the Modern Identity

CHARLES TAYLOR

"Surely one of the most important philosophical works of the last quarter of a century."
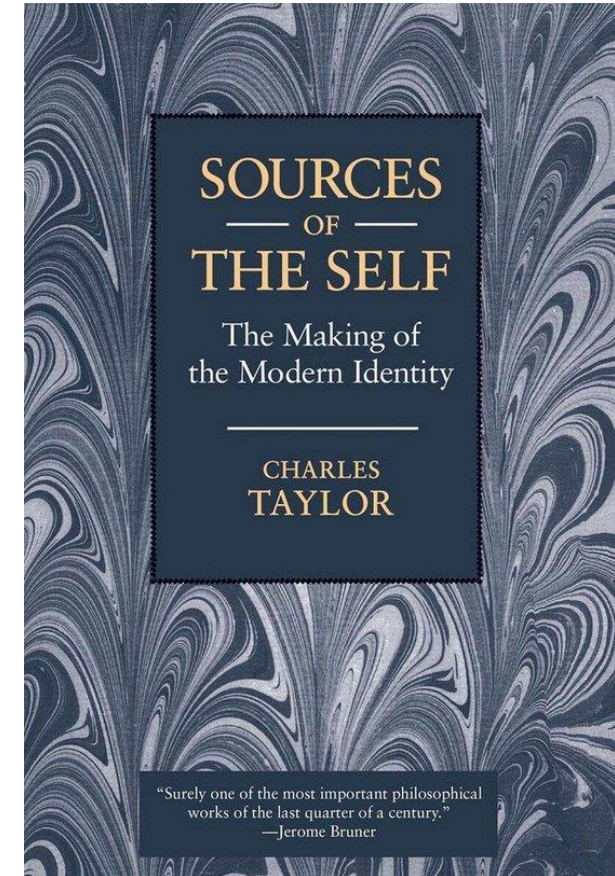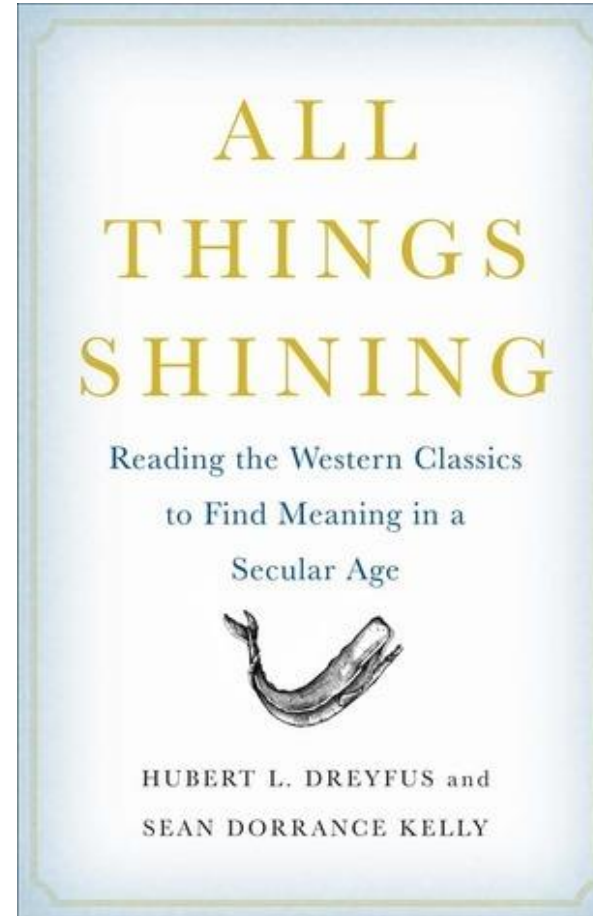—Jerome Bruner

The "enchanted" medieval world, with a "hierarchical complementarity" (Taylor) between microcosm and macrocosm.

There are seven windows given to animals in the domicile of the head, through which the air is admitted to the tabernacle of the body, to enlighten, to warm and to nourish it. What are these parts of the *microcosmos?* Two nostrils, two eyes, two ears, and a mouth. So in the heavens, as in a *macrocosmos*, there are two favourable stars, two unpropitious, two luminaries, and Mercury undecided and indifferent. From this and from many other similarities in nature, such as the seven metals, etc., which it were tedious to enumerate, we gather that the number of planets is necessarily seven.[1]

[1] Quoted from S. Warhaft (Ed.) *Francis Bacon: A selection of his works*, Toronto, 1965, p. 17.

4

- Descartes' self-sufficient power to doubt everything but himself and to guarantee his knowledge (<mark>DISENCHANTMENT & VALUE</mark>)

- Kant's self-sufficient power to give himself the moral law

- Hegel's freedom grounded in mutual recognition of equal others ["Freedom to salute the police" (B. Russell)]

- Humboldt & Mill: Cultivating your unique individual inspirations and talents

- Nietzsche & Weber - Freedom to create ("posit") own's own values

- Modern technologists: Powerfully solving all problems, achieving "technological maturity" in a "solved world," maximizing utility, optimizing all options.

ALL THINGS SHINING

Reading the Western Classics to Find Meaning in a Secular Age

HUBERT L. DREYFUS and SEAN DORRANCE KELLY

SOURCES OF THE SELF
The Making of the Modern Identity

CHARLES TAYLOR

"Surely one of the most important philosophical works of the last quarter of a century."
—Jerome Bruner

Heidegger on being human: "Dasein"

Dasein is the entity "whose own being is an *issue* for it."

These various interpretations of human well-being and autonomy are various ways the structure of Dasein's caring can be filled out.

This is a history of interpretations of what is worthy of our care.

We "disclose the world" as a field of mattering.

We are self-interpreting beings, whose nature is defined by our care.

SEIN UND ZEIT

VON

MARTIN HEIDEGGER

MAX NIEMEYER VERLAG TÜBINGEN
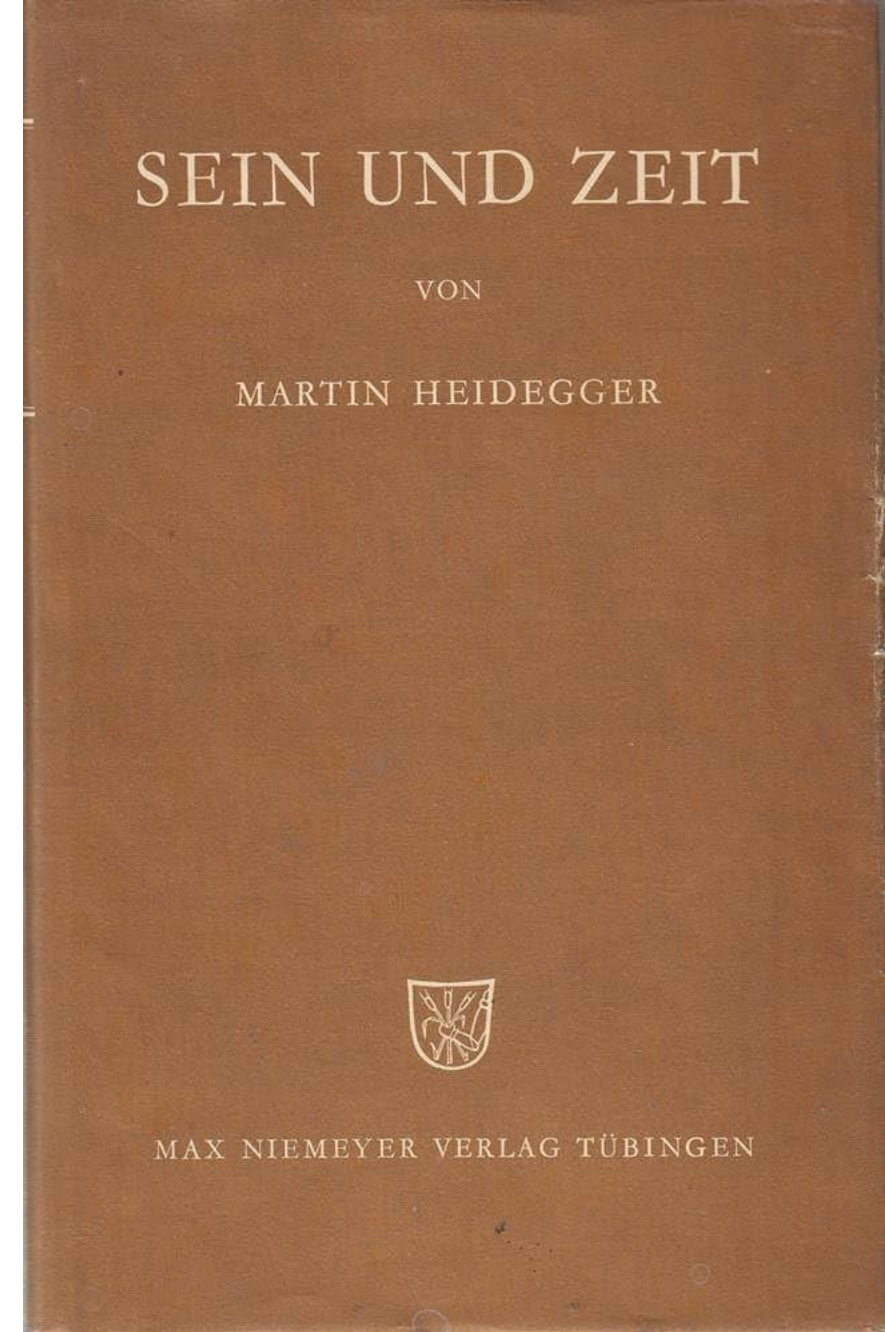
Being and Time (1927)

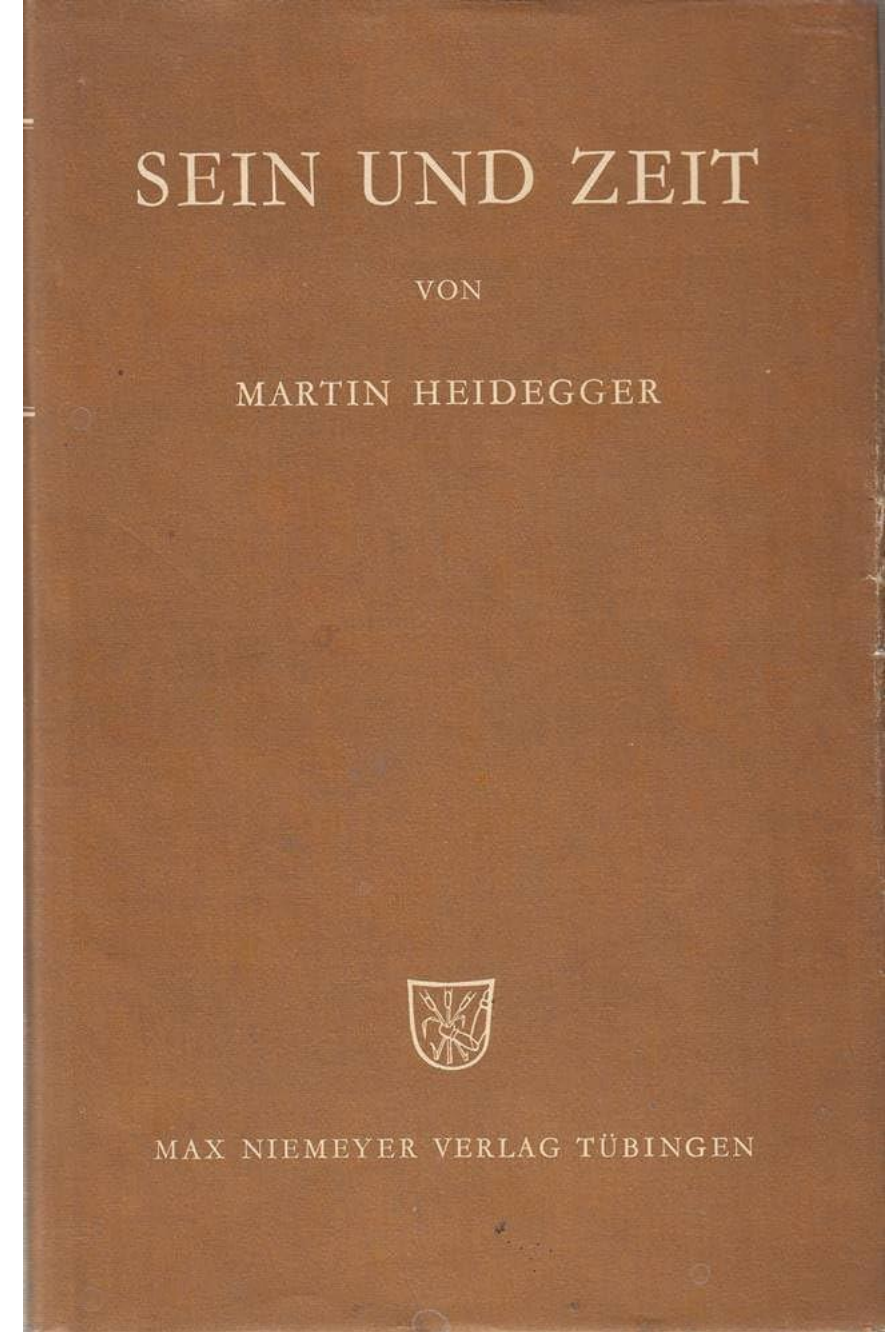Our "caring" way of being is such that it can misunderstand itself.

An ontology of "values" covers up and misconstrues this structure of care.
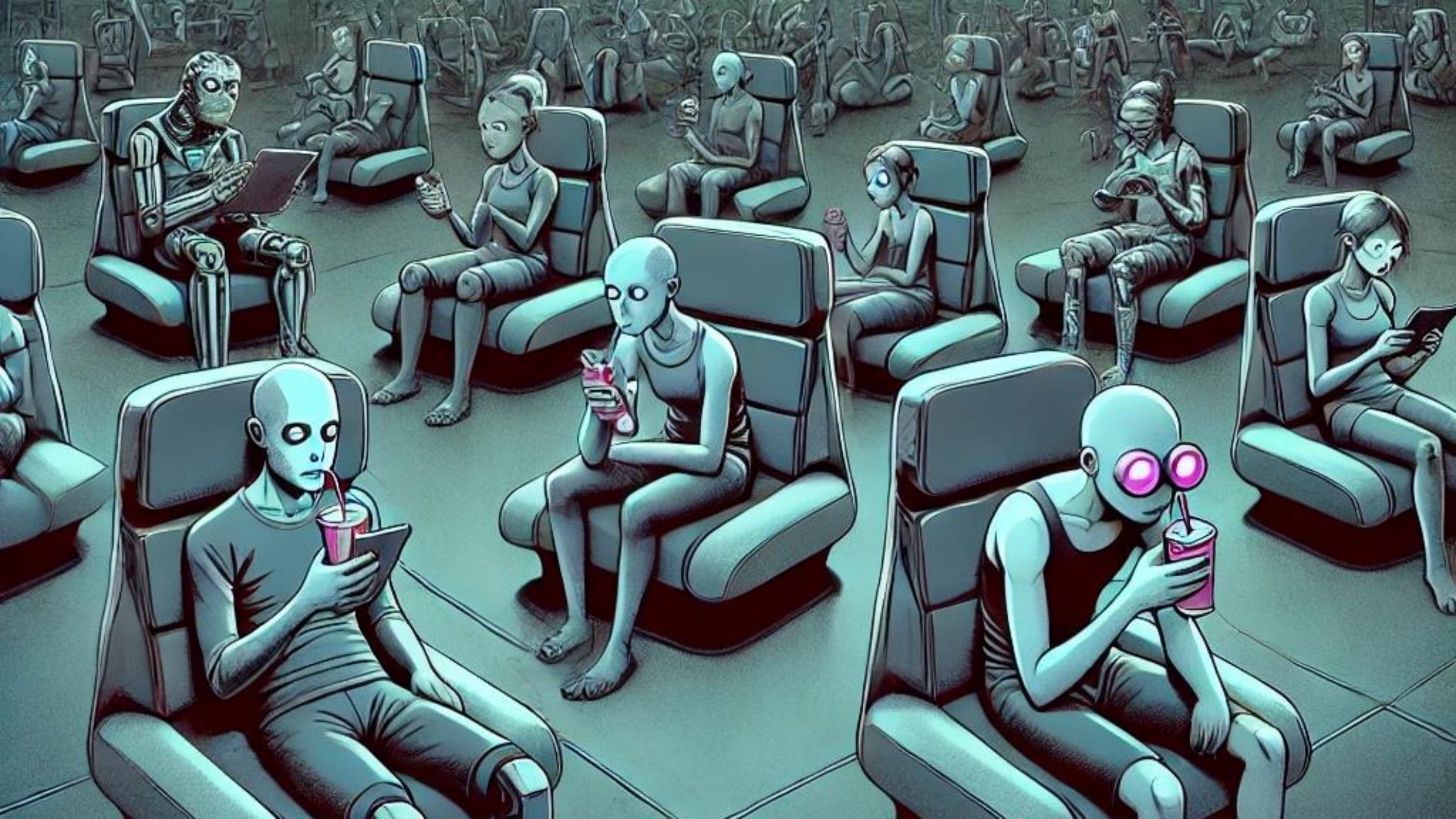
"Values" are an ontological self-misunderstanding

Our caring is such that it can be usurped, diluted, offloaded in our social practices.
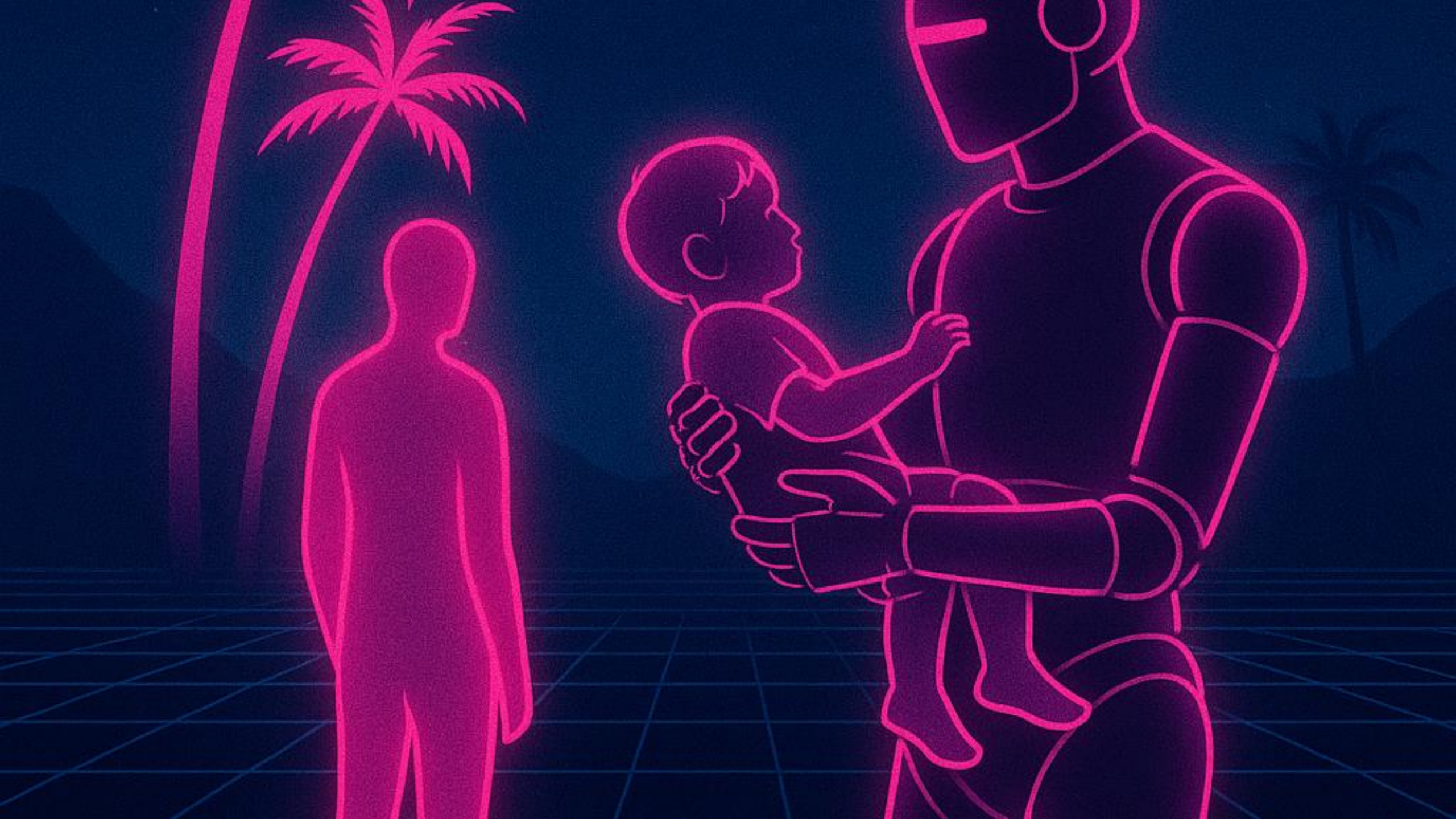
An ontology of values makes this more likely.

→ The *real* problem of human enfeeblement.

Being and Time (1927)

Conceding power to AIs could lead to human enfeeblement. Even if we ensure that the many unemployed humans are provided for, we may find ourselves completely reliant on AIs. This would likely emerge not from a violent coup by AIs, but from a gradual slide into dependence. As society's challenges become ever more complex and fast-paced, and as AIs become ever more intelligent and quick-thinking, we may forfeit more and more functions to them out of convenience. In such a state, the only feasible solution to the complexities and challenges compounded by AIs may be to rely even more heavily on AIs. This gradual process could eventually lead to the delegation of nearly all intellectual, and eventually physical, labor to AIs. In such a world, people might have few incentives to gain knowledge and cultivate skills, potentially leading to a state of enfeeblement [52]. Having lost our know-how and our understanding of how civilization works, we would become completely dependent on AIs, a scenario not unlike the one depicted in the film WALL-E. In such a state, humanity is not flourishing and is no longer in effective control.

Dan Hendrycks, Center for AI Safety: *Introduction to AI Safety, Ethics, & Society* (p.32)
Cf. Stuart Russell, *Human Compatible* (p.255).

A Turning Tide? Towards a greater acknowledgement of *care*:

- Brian Christian, "Computational Frameworks for Human Care" (2025)
- Leonardo Christov-Moore, et al, "The Contingencies of Physical Embodiment Make Care Possible" (forthcoming)
- Thomas Doctor, et al, "Biology, Buddhism, and AI: Care as the Driver of Intelligence" (2022)
- Joe Edelman, "Values, Preferences, Meaningful Choice" (2022)
- Oliver Klingefjord, et al, "What are human values, and how do we align AI to them?" (2024)
- Joel Lehman, "Machine Love" (2023)
- B. Scot Rousse and David Spivak, "The Angst and Value of Accounting for Care" (forthcoming)
- Olaf Wittowski, et al, "Toward an ethics of autopoietic technology: Stress, care, and intelligence" (2023)