AI and Humanity

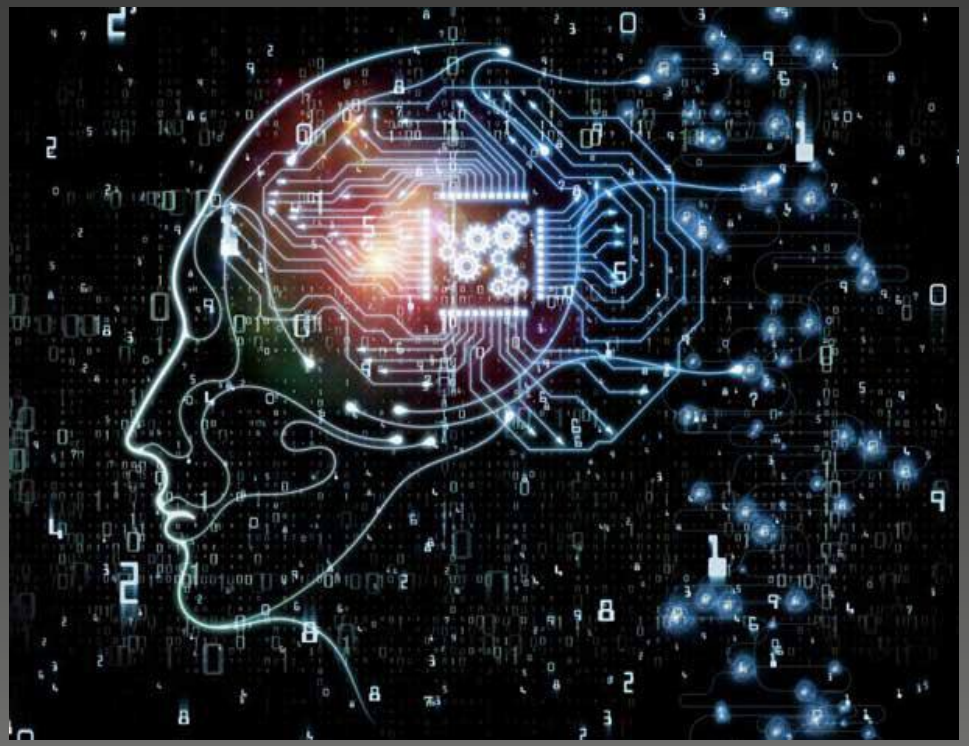
"Intelligent Technology and the Attention Economy: A Buddhist Perspective on the Risks of Consciousness Hacking"

Simons Center for the Theory of Computing

Peter D. Hershock, East-West Center

A “Cambrian explosion” is underway in the synthesis of purpose-generating, carbon-based human intelligence and purpose-implementing, silicon-based machine intelligence.

A 4th Industrial Revolution
but one that is:
as *ontological*
as it is industrial
as *ethical*
as it is technical



The human risks and rewards of Intelligent Technology cannot be taken into full account without taking a relational turn, both ontologically and ethically.

- ethical engagement with AI is hampered by presuming ontologically individual moral agents, actions, and patients
- the full risks of AI are obscured by failing to distinguish critically between tools and technologies
- the greatest threat posed by intelligent technology
  - is not a *technological singularity* it is an *ethical singularity*
  - a collapse of the opportunity space for freely human course correction triggered ironically by choice-mediated attention exploitation and consciousness hacking

Buddhist Nonduality: A Relational Turn

All things occur interdependently

- all things *are* what the *mean* to/for each other

There are two kinds of causality and time

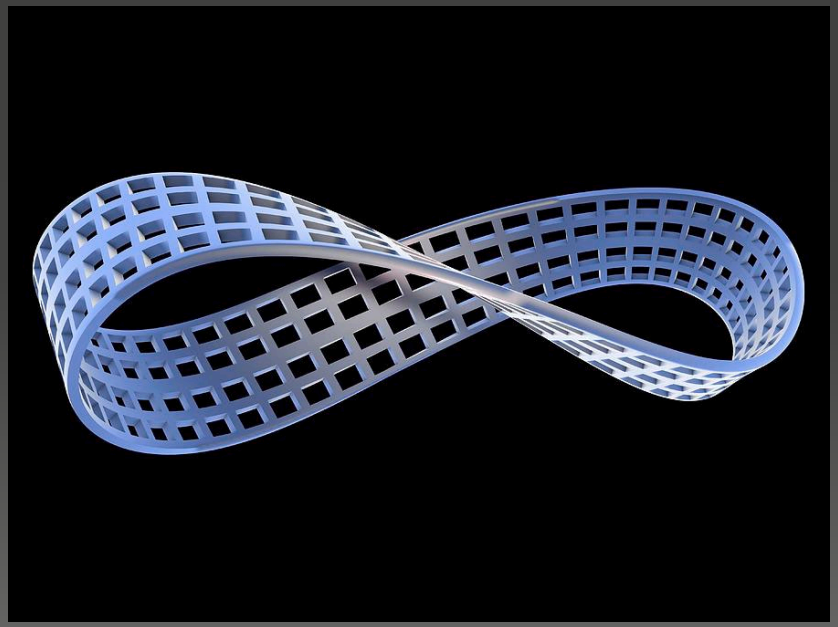
- linear sequence-ordered temporality
- nonlinear significance-ordered temporality

Karma:
the diffractive interplay of these causalities and temporalities

Buddhist Nonduality: A Relational Turn

Taken together, these insights entail the nonduality of the ontological and axiological realms—that is, the realms of matter and what matters, or the physical and the phenomenal

This compels rethinking consciousness and the human-technology-world relation



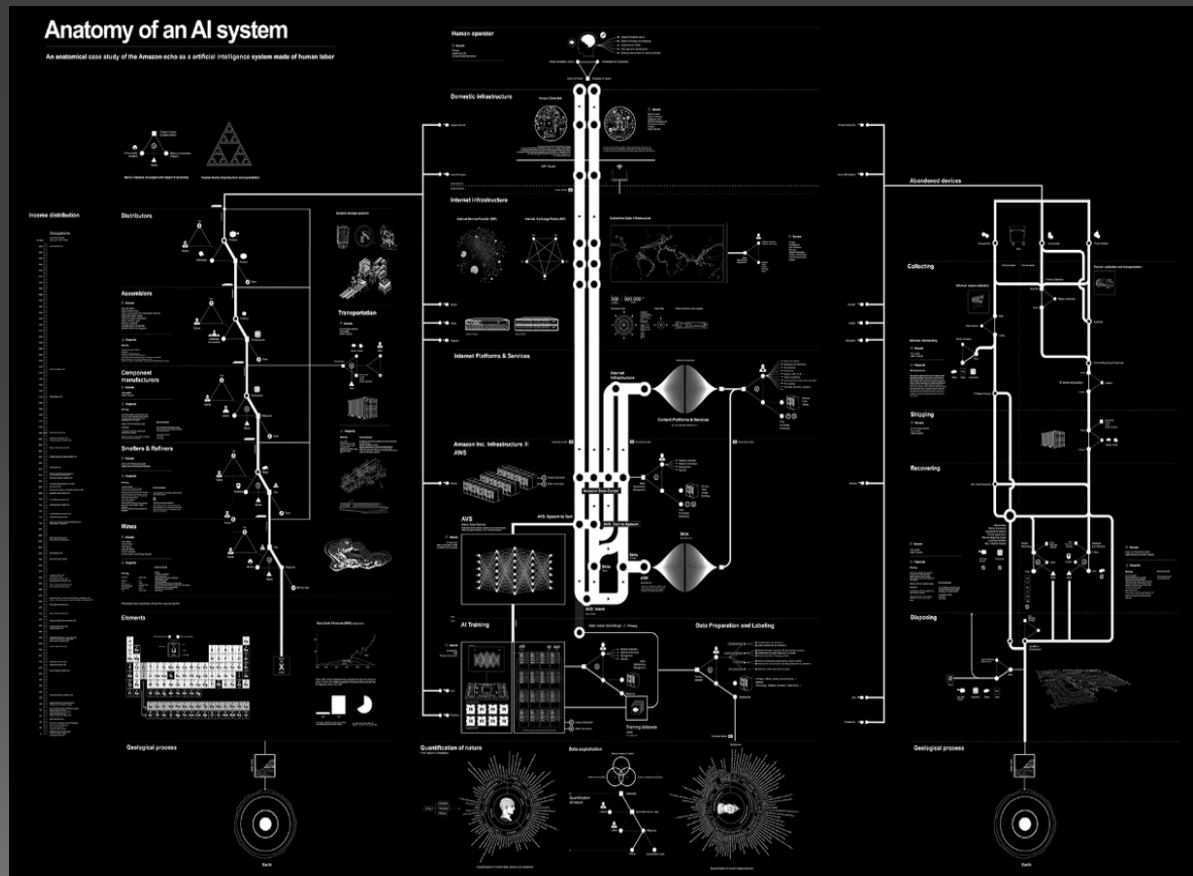
The Irreducibility of Technologies to Tools

Seen relationally, technologies and tools occupy distinct ontological and axiological registers.

Tools: localizable *things* that extend /augment action.

Technologies: non-localizable *relational systems* that transform what we do, how, and why.

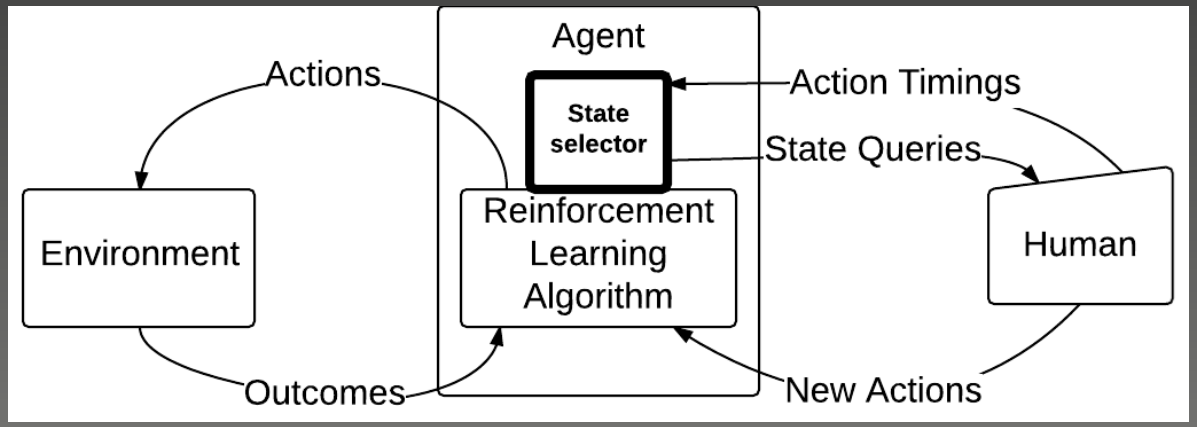
We do not *build/use* technologies. We *participate* in them.



Humans **In** the Loop

Humans in the loop: inserting a human between the observing, orienting, and decision-making phases of machine-environment interactions, and the decision-enacting and world-affecting phase.

A sound precaution to limit the autonomy of machine agencies that works to the precise extent that the human in the loop retains *evaluative independence* and *freedom of intention* in an essentially external user-tool relationship.



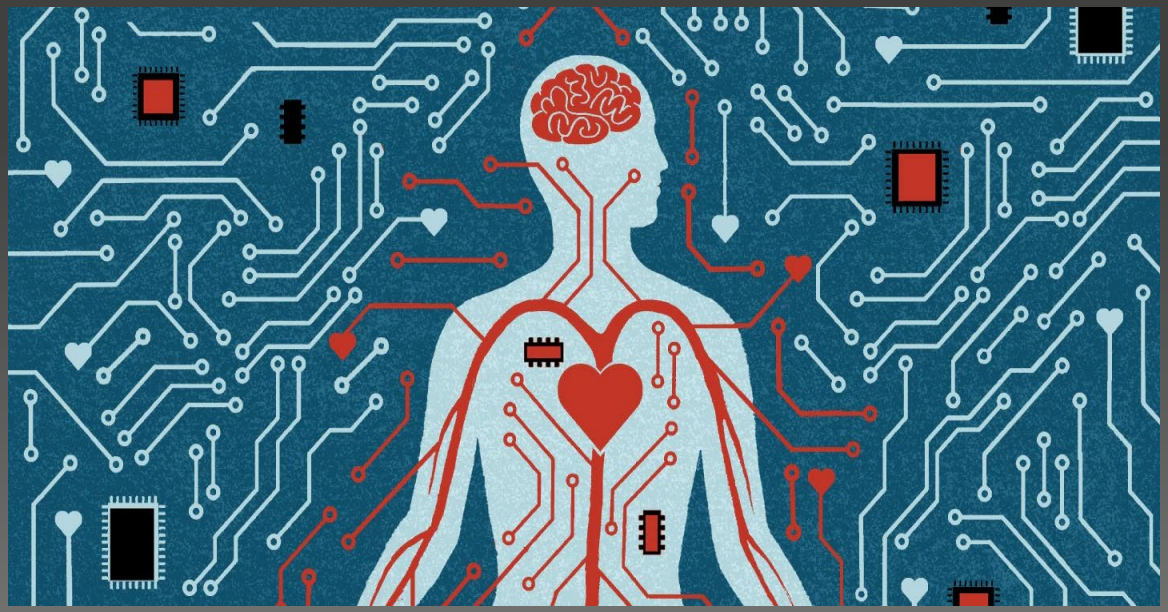
Humans As the Loop

The conceit of independence is not tenable when humans cease being in the loop of *smart tool* operations and become functionally integral loops within *intelligent technology* as a dynamically evolving relational medium.

Humans as:

consumers of individually-targeted goods and services

producers of training data for systems accelerating and expanding processes of *attention capture* and *exploitation*



The New Attention Economy

If we are presumed to participate in intelligent technology as ontologically independent individuals, this double-duty can be framed as a free exchange or as a predatory labor practice.

Either way, ethical concerns remain technically tractable

But if we are relationally constituted, the ethical ramifications of intelligent technology extend beyond data privacy and ownership, decision-making accountability and transparency, and inequitably distributed risks and rewards.

The New Attention Economy

As an ironic consequence of expanding individual freedoms of digital choice, significantly human autonomy is increasingly at risk.

- as ML systems yield unprecedented *epistemic powers* to predict human behaviors, desires, and fears
- they yield no less unprecedented *ontic powers* to produce behaviors and beliefs, affecting how and how freely we relate

We risk forfeiting our most basic human rights and freedoms: our rights to freedom-of-attention and freedom-of-intention

Consciousness Matters

From a Buddhist perspective, the digital manipulation of attention amounts to mass experimentation in the hacking of human consciousness.

Current Western theorizing about human consciousness can be sorted into four distinct streams:

- reductionist (materialist or idealist): consciousness is reducible to neural activity
- functionalist: consciousness is software running on an organic or perhaps inorganic substrate
- dualist: physical and phenomenal realities interact, but are ontologically distinct
- enactivist: consciousness originates in body-environment sensorimotor coupling

Consciousness Matters

Buddhism theorizes consciousness as irreducibly relational. Rather than being a property of an organism or arising through the coupling of an organism and its environment, consciousness consists in the *coherent differentiation* of sensing and sensed presences.

- brain-body-environment relations are elaborations of the *infrastructure* of consciousness
- *materializations* of what consciousness *does*

Ultimately, consciousness consists in differentiations of matter and what matters that are conducive to the continued, improvisational elaboration of potentials for coherence.

Evolution is consciousness mattering

Consciousness Matters

According to this relational theorizing, organic consciousness is not singular, it is sense specific. Humans comprise relations occurring within/among eight distinct kinds of consciousness:

- visual consciousness consists in elaborating visual relations and order, and so on for hearing, smell, taste, and touch
- a sixth, cognitive consciousness elaborates relations among the dynamics occurring within and among the other five senses

With significant implications for machine consciousness, there is not necessarily anything "it is like" to see, hear, taste, touch, smell, or think.

Consciousness Matters

Persistent experiential presence and subjective identity involves the work of two other consciousnesses.

- a subjective 7th consciousness that elaborates and evaluates relations among sequentially ordered sensory currents (the work of the 6th consciousness) and significance ordered patterns of events (the work of the 8th consciousness)

Subjective presence consists in the diffractive interplay of material/phenomenal occurrences in two distinct and yet interpenetrating temporal dimensions.

- the advent of the 7th and 8th consciousnesses marks a shift from *responsive* to fully *responsible* presence—the advent of recursively intentional coherence and differentiation

The Neuro-Ecological Compass of Consciousness

Some recent neuroscientific support for Buddhism's relational theorizing of consciousness:

- valued psychedelic and meditation experiences correlate with increased brain entropy
  - a "disordering" of the default mode neural network (the neuro-linguistic infrastructure of the 7th consciousness)
  - dissolving habits of thought, feeling, and action that ordinarily constrain attention and limit responsiveness
  - consciousness minimizes relational entropy, enhancing ecological differentiation and coherence
- inter-brain neural entanglement is affected by what we think or expect to be true
  - brains are biosocial, relational media, and consciousness can be significantly shared

Expanding the Ethical Compass in Response to Intelligent Technology

These findings challenge reductionist convictions of mind-body relations. But they also compel significantly expanding the ethical compass in response to intelligent technology.

- using machine learning tools to predict and produce human behaviors is ethically analogous to inserting brain electrodes
- algorithmic curation of content in digital environments affects social media experiences in ways analogous to altering the biochemical environment of the brain

This is nothing less than mass experiments in hacking the infrastructure of human consciousness.

Expanding the Ethical Compass in Response to Intelligent Technology

Might the objective function of machine learning systems be to foster compassionate and collaborative relational dynamics? Might humans be algorithmically engineered to be ethical by design?

- past eugenics projects and questions about whose ethics get baked into the infrastructure of consciousness should be caution enough

But in any scenario, it would mean the end of *human consciousness mattering*

- without freedom-of-attention, there can be no freedom-of-intention
- without freedom-of-intention, we would be unable to transform who we *are* by transforming what we *mean* both to and for one another, unable to engage in the most important of the collaborative and improvisational human arts—the ethical art of fully responsible course correction