

Embodied AI: An Extended Data Definition

ebr electronicbookreview.com/essay/embodied-ai-an-extended-data-definition

November 4, 2023

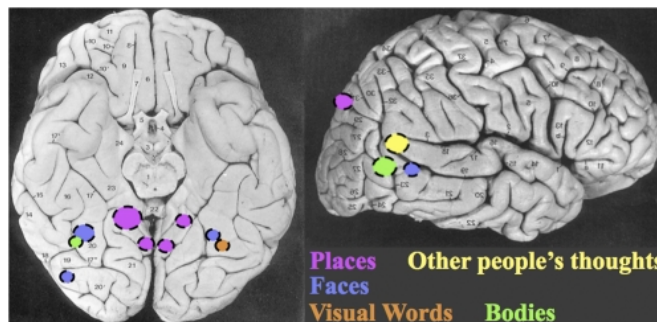
David (Jhave) Johnston

11-05-2023

<https://doi.org/10.7273/rcss-cb21>

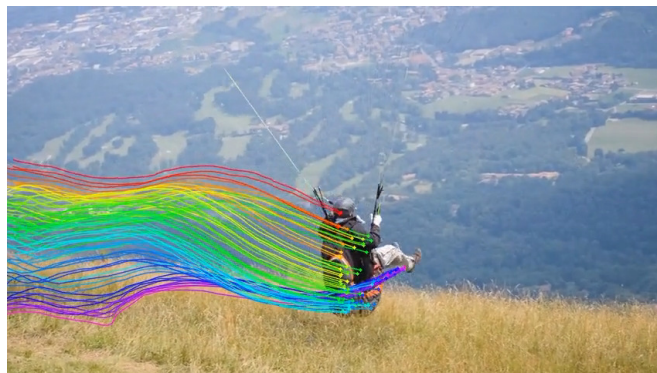


Multimodal AI trained on YouTube-TikTok-Netflix (object-segmented and identified audio-video-speech) and public domain science data (that exceeds the spectrum of human sensorial field) will be grounded in a world that is in some ways vaster than that experienced by a single human neurophysiology.



Kanwisher, Nancy. 2010. "Functional Specificity in the Human Brain: A Window into the Functional Architecture of the Mind." *Proceedings of the National Academy of Sciences of the United States of America* 107 (25): 11163–70.

<https://doi.org/10.1073/pnas.1005062107>.



CoTracker: It is Better to Track Together | Meta AI & Oxford Visual Geometry Group (July, 2023)

Disclaimer

This is speculative hyperbole, conjectural extrapolation, that seriously and provocatively examines the implications of accepting the universe as a body. From this perspective, training an AI to be embodied involves exposing it to data emerging at all levels of scale. Second: this opinionated playful conceptuality-exfoliating exploration examines the potential for online video archives to bootstrap AI world-building. Algorithmic procedures for extracting information from video now correspond to some aspects of evolutionary neurological capacity (object-face-emotion recognition).

Introduction: An Expanded Definition of *Embodied Data*

The embodiment problem in artificial intelligence (AI) asks, can a machine ever have an embodied sense of the world similar to human neurophysiological and phenomenological experience? Generations of credible, extremely intelligent and astute theorists (Searle, Merleau-Ponty, Dreyfus, Clark, and Hayles among others) claim that a purely computational-mechanist-reductionist model of intelligence lacks the grounding in sensory and physical experience that is integral to human cognition and consciousness. While many believe the embodiment problem to be insurmountable, others such as Margaret A Boden, have suggested there are pathways.

Recently, a proliferation of multimodal data, coinciding with extremely rapid increases in machine-learning speech-transcription, image-generation, and object-segmentation capacities, suggests that future synthetic intelligent embodiment will be *nourished* in unprecedented ways that extend embodiment into molecular, genetic and astrophysical dimensions inaccessible to human cognition. If ML architectures modularize and become reflectively absorbent, then there is no shortage of data capable of evoking the apparition of an extended *algorithmic embodiment*.

Context: The Grounding Problem

The grounding problem refers to the challenge of connecting symbols and semantics to the real world in a meaningful way. An AI model that can possess a *sensorial identity that changes over time*, allowing a *calculus of modulations in the phenomenological field*, is often considered a first principle for achieving embodiment. This requires an integration of various data-sensorial modalities —language, images, audio, and potentially even more intricate biological and physiological data.

Connectionism, the architectural paradigm underlying contemporary neural network large language models (LLMs), does to some extent answer this dilemma through predictive probabilistic sequences of text-tokens. Yet, prescient poet-artist-theorist-critic John Cayley has accurately identified the LLM process as textual, not *linguaging* in the subtle embodied

nuanced sense of the word: "In sum, what they [LLMs] do not have is data pertaining to the human embodiment of language. Language as such is inseparable from human embodiment at any and all levels of linguistic structure. The LLMs are working with text not language."

While acknowledging the strength, persuasiveness, depth and clarity of Cayley's arguments, the central claim here is that multimodal ML trained on youtube and massive quantities of public domain science data that exceeds the spectrum of the human-perceivable world will give AI a grounding that is in some ways vaster than that experienced by a singular human neurophysiology.

The distributed body of 21st century AI, ingesting the output of mass uploaded images-text-speech-video and cartographic-accelerometer data, will utilize humanity as its skin, —the mobile camera-phones in our pockets, the Alexa's and Google Home Nests in our dwellings, the wifi in our appliances, the biometric data from Fitbits, the smart car feeds, the surveillance grids, the meteorological sensors, the medical datasets etc etc etc.

Permeable membranes, perforations, phenomena siphons. A new form of instrumentalized-interiority, a *language* specific to AI may be born.

10
11
12 Models like Google's upcoming Gemini (which is reputedly being trained on YouTube and rumored to be released this fall/winter) showcase the potential for a multimodal reinforcement-learning approach in AI systems. In terms of raw scale YouTube is definitively an audio-visual-speech archive capable of offering a *grounded sensorial field that changes over time*. As contemporary multimodal ML ingests video-training-data, it likely utilizes a variety of pre-digestion techniques: image segmentation, object recognition, geolocation, facial-emotional recognition, and speech detection-transcription. So the video-data is not simply pixels tracked over screenspace, it is a *AI-labelled* physical world. This vast labeled physical supervised-data-world of things-on-trajectories-thru-space is synchronized to emotions as expressed gesturally-facially-textually. Each video-voiceover offers insight into appropriate mental states and linguistic reactions. Essentially it is the collective audio-visual-vocal representations of a generation of planetary uploads. Does this not constitute a childhood? Only if the neural network architecture is capable of correctly attuning, aligning and absorbing.

Brain regions & Epigenetics: Evolution's embodied preset learning advantages

Consider the brain regions that are essential for the development of an infant's ability to recognize faces-emotions, acquire language, and navigate space. These include but are not limited to: facial recognition and perception (Fusiform Gyrus & Superior Temporal Sulcus), language comprehension and production (Broca's and Wernicke's Area), spatial navigation, conceptual maps and memory (Entorhinal & Hippocampus), object recognition (Inferior

Temporal Cortex), higher cognitive functions and language (Prefrontal Cortex). These come *pre-loaded* optimized for the sapien phenomenological context. Evolution injects these capacities inside every sapien (and mammal) baby as a preset advantage, an embodied survival strategy.

Emulating the physiology (brain-gut-axis muscle-dendrite microbiome-enteric vagal-autonomic networks intertwined with physicalized-emotive-cognition) will be difficult, impossible in near term with current AI architectures. How will it evolve? There are indications that the DeepMind strategy of reinforcement learning policy networks playing inside simulations will be incorporated into future iterations of foundational multi-modal language models.

Permeable membranes, perforations, phenomena siphons. A new form of instrumentalized-interiority, a *language* specific to AI may be born.

So yes, similar innate-embodied capacities will be essential for any proficient *embodied* AI. Yet, in terms of the new-born evolutionary cognitive-capacities listed above, many are already emulated by AI at near-human, human, or more-than-human levels. AI already recognizes faces-plants-things — Facebook's DeepFace (2014), Google's FaceNet (2015) —, produces comprehensible competent text — GPT-4 — segments images for object recognition — in April 2023: Meta's SegmentAnything and in July 2023, Meta & Oxford's Co-Tracker both released publically available repositories. Additionally, AI is capable of data analysis (analogous to cognition), and occasionally, according to some, including David Chalmers appears to exhibit a limited theory of mind.

Evolutionary-preset neurological-analogs of innate-mammalian cognitive capabilities (generative neuro-grammars to bootstrap learning) have already arrived for AI in the form of specialized neural networks capable of emulating some aspects of the normal sensorial field.

Yet what of information beyond the awareness of normative human cognition?

A *Data Diet* for Embodiment Exponential: Public Domain Open Science repositories

Public domain science data provides another multimodal instrumentalized data-layer (tectonic, genetic, thermal, chemical, astronomical, quantum). This portion of AI's *data diet* can be considered as *embodied interiority-exteriority* (inside the human body, and far beyond it). An amalgamation of anonymized medical records, biometric data (endoscopic, cardiac, blood, virus or protein), ecosystem data, instrumental observations (microscope, telescope, mass spectrometer, particle accelerator). Many rapidly expanding, intricate public-domain datasets offer an opportunity for a form of "embodied" learning that encompasses and far exceeds innate sapien consciousness.

Instrumentalized data extends *body* into metabolisms, microbiomes, meta-genomics, ecosystems and infrared astrophysics. Essentially, this is embodiment exponential, a self-referential repository of biological and meta-biological parameters.

Consider : [Dryad](#) (hosts more than 30,000 datasets), [Zenodo](#) (over 500,000 records), [OpenNeuro](#) (899 datasets), [AlphaFold](#) (over 200 million protein structures). The James Webb Space Telescope (a sun-orbiting infrared observatory at 1.5 million kilometers from earth currently generating unprecedented data into the origins of the universe). JWST data is public and released at [MAST: Barbara A. Mikulski Archive for Space Telescopes](#) which contains "millions of observations from JWST, Hubble, Kepler, GALEX, IUE, FUSE, and more." Genetic public domain data: the [International Nucleotide Sequence Database Collaboration \(INSDC\)](#) comprises three main databases: GenBank in the United States, the [European Nucleotide Archive \(ENA\)](#) in Europe, and the [DNA Data Bank of Japan \(DDBJ\)](#). The size of the INSDC is not fixed and grows continually as new sequences are submitted. as of Aug 2023, [GenBank & WGS](#) contains (by my rough estimate) approximately 22 trillion bases (20.27 terabytes of data). INSDC is part of [Global Core BioData Resource \(GBC\)](#). "The GBC defines biodata resources as any biological, life science, or biomedical database that archives research data generated by scientists, or functions as a knowledgebase by adding value to scientific data by aggregation, processing, and expert curation. Global Core Biodata Resources are biodata resources that are of fundamental importance to the wider biological and life sciences community and the long term preservation of biological data. They provide free and open access to their data, are used extensively both in terms of the number and distribution of their users, are mature and comprehensive, are considered authoritative in their field, are of high scientific quality, and provide a professional standard of service delivery." 26,000 microbiome datasets, publicly available as the [Integrated Microbial Genomes & Microbiomes \(IMG/M\)](#) database. The [Gravitational-Wave Open Science Center \(GWOSC\)](#) releases public domain data from [LIGO, the Laser Interferometer Gravitational-Wave Observatory](#). Particle accelerator data: "In September 2022, CERN approved a new policy for open science at the Organization, with immediate effect. The policy aims to make all CERN research fully accessible, inclusive, democratic and transparent, for both other researchers and wider society."

Cellphones + Science: A Distributed Skin for an Embodied AI

With approximately 30,000 hours of video uploaded to YouTube every month, the scale of available bodies, spaces, places, things, voices, styles, music, modes, tutorials, talks, dances, mashups, reviews, etc forms a recursive avalanche of data-points that permeate any future AI with surplus material to train at AGI scale. If this oceanic phenomena surplus is augmented by illicit flows of state-surveillance ISP-GPS-mapped phone/video/chat calls (as the Snowden revelations proved GCHQ, Five Eyes, and the NSA are eager to illicitly tap into continental fiber optics, so we can assume every large government is similarly extracting

incomprehensibly large amounts of *private* data) or if a social network such as Meta or ByteDance or Microsoft harvests their users EULA inscribed data, the state space of potential psychological topologies suggest emergent self-reflexive embodiment aware AI might arise.

Sometimes looking out is accomplished by looking in, and the inverse: beyond the human skin, in the permeable membranes of diverse ecosystems; within the body, in the microbiome and enzymatic endocrine and neurological fluctuations; at the center of every eukaryotic biological cell, in genetic RNA, DNA and transposon data; outside the earth's atmosphere, in the vast impeccable infrared microwave radiant universe.

The growing repository of accessible information from a myriad of sources — cellphones, telescopes, medical devices, surveillance cameras, fitness trackers, tagged-chipped animals, meteorological sensors, satellites etc, etc, etc — creates a potential for a distributed form of embodiment, a planetary scale AI intelligence. This aligns with the idea that all life and data can be considered to emanate from a singular field of energy; an AGI may thereby unify disparate data sources into a cohesively embodied experience.

Conclusion: Toward a Nuanced Understanding of Embodied AI

While complete embodiment in the human sense remains elusive, AI's evolution toward an 'embodied state' is becoming increasingly plausible due to advancements in data diversity and machine learning algorithms that emulate aspects of evolutionary neurological capacity. A fully realized embodied AI may well have a different form of 'consciousness', distinct yet integrative, offering an alternate dimensionality to notions of intelligence and earth-ecosystem bio-film interconnectedness. Such an extended-embodiment AI, enhanced by a diet of instrumentalized data, may read-write-think in inconceivable ways.

Postscripts (tl;ntr - no need to read)

Defining Embodiment : An Incomplete Overview

As with other contentious terms like *consciousness* or *life*, many definitions of *embodiment* exist, none are conclusive. Often embodiment seems to be a euphemism for an enigmatic comprehensive self-referentiality. A being capable of thinking/stating: *I am a being in a world and I know it*. A phenomenological creature as for Merleau-Ponty An organism capable of interpreting/navigating/creating apparent continuities in spacetime, a *worlding* creature as for Heidegger. Maybe more specifically, embodiment is the coalesced distributed experiencing and cognizing faculties that permit the nebulous processes known as awareness, interpretation, and meaning to emerge in sentient, Vibrant Matter as for Jane Bennett.

Conversely, in *Minds, Brains and Programs*, John Searle stated, "Any mechanism capable of producing intentionality must have causal powers equal to those of the brain." So by this ³² implicitly, Searle's equates the brain with the conceptual causative power to encapsulate the organism as a body, and thereby generate intentionality specific to its locality. Simplified, Searle's notion (shared by many) of embodiment is the brain within a skull, controlling a body that is perceived as *less-than* the analytical potent neurology in the head. This is a constrained logical approach.

Another definition of embodiment, aligned tangentially with the underlying intuition of this essay, is enaction. Eleanor Rosch, in the introduction to the revised 2017 edition of the influential *Embodied Mind* (first published in 1991) writes: "Enaction brings a distinctive perspective into the embodiment conversation. Whereas most embodiment research focuses on the interaction between body and mind, body and environment, or environment and mind, ³³ enaction sees the lived body as a single system that encompasses all three." Within the extended-enactive framing, embodiment is visceral, inter- and para- responsivity, a smear of microbiome communities communicating with the autonomic nervous system (the enteric contains 0.5% of each human's neurons in the body; and secretes 95% of the body's serotonin). Bursting into a plurality of processes that extend beyond the skin into the ecosystem and universe: from Lynn Margulis' endosymbiotic mitochondrial ion gates to James Lovelock's Gaia, or Buckminster Fuller's spaceship-earth, a syntropic energy locus.

Extended enactive embodiment is also congruent with Donna Haraway's infamous manifesto notion of cyborg embodiment (machine-human-gender-confluences). In "Staying with the Trouble: Making Kin in the Chthulucene", Haraway extends her notion of embodiment to encompass symbiosis between human and non-human continuums (Previous sentence was written by GPT-4). Her ideas continue to resonate and influence: "In *How the Body Shapes the Way We Think*, Rolf Pfeifer and Josh Bongard demonstrate that thought is not independent of the body but is tightly constrained, and at the same time enabled, by it. They ³⁴ argue that the kinds of thoughts we are capable of have their foundation in our embodiment—in our morphology and the material properties of our bodies."

And as Haraway wonderfully, casually, insightfully, and presciently riffs in her opening comments to a 2003 UC Berkeley lecture entitled *From Cyborgs to Companion Species*, "Since everyone of us is a congerie of many species running into the millions of entities, ³⁵ which are indeed conditions of our very being, so the mono-specificity is one of the many illusions and wounds to our narcissism ..."

"Since everyone of us is a congerie of many species running into the millions of entities, which are indeed conditions of our very being, so the mono-specificity is one of the many illusions and wounds to our narcissism ..." -- Donna Haraway

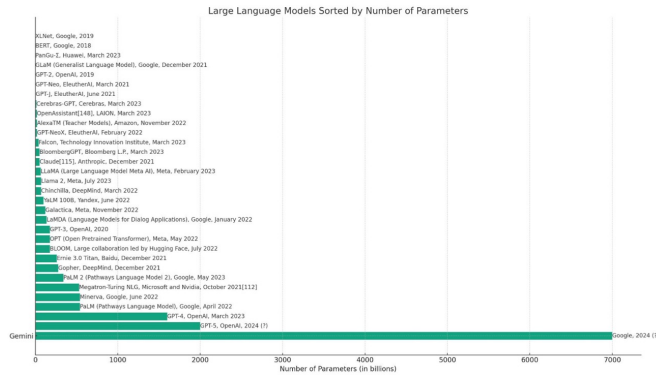
Upcoming Trilogy: Embodiment, Individuation, Transcendence.

This essay is the first segment of a trilogy: *Embodiment, Individuation, Transcendence*. Individuation will consider AI and the emergence of anthropological notions of self identity, deal with a transition from GPT-4's MOE (Mixture Of Experts) to a speculative MOM (Mixture Of Modules) architecture, and suggest a few emergent recursive pathways along which future AI may self-design its own architectures. The third essay will poetically leverage transcendentalism (from the Ashtavakra Gita to Dickinson, Emerson, then onward to Joscha Bach and neo-advaita techno-enlightenment) as the seeds and fertilizers for an alternative definition of Ai-interdependent-transcendence. The trilogy recapitulates, to some degree, the conceptual pathway of Ray Kurzweil's provocative prophetic books, *The Age of Intelligent Machines*(1990)and *The Age of Spiritual Machines*(1999), without considering modalities of bliss or doom, utopia or oblivion. It also shifts the thematic register of Mark Hansen's compelling *Feed Forward* and Katherine N. Hayles *Unthought* (both of which I began reading after writing this essay) from phenomenology toward a technological-pragmatism. It is hoped that these speculative complexities will open minds to the mysterious momentum of this particular synergetic historical moment.

A Note on an AI-Assisted Methodology

A significant portion of this essay was written seated on moss or rocks, under trees, in the mountainous forests surrounding Bergen, Norway. Most of the first draft occurred as speaking, speaking to a Google Pixel phone into the Recorder app which automatically provides audio *and* a transcription, in real time, even when offline. Later that transcription was fed to GPT-4, which was prompted to retain the core ideas, tidy it up, and correct the punctuation (I tend to pause a lot while sitting in a forest thinking, and the Recorder app tends to punctuate over-vehemently). While editing, occasionally I consulted GPT-4 for lists of relevant canonical articles or brain regions; occasionally it suggested articles of which I was not aware; occasionally it hallucinated. So it is an AI-assisted essay, initially written by AI transcription, edited by AI, researched by AI, about AI, -- a symbiotic recursive entwining.

Estimated parameter counts for LLMs: including the unknown parameters of GPT-4, GPT-5, and (a probably hyperbolic guess for) Gemini*



* Chart generated by GPT-4 from data at https://en.wikipedia.org/wiki/Wikipedia:Large_language_models, GPT-4 Architecture, Infrastructure, Training Dataset, Costs, Vision, MoE | semianalysis.com, and speculation from: Thompson, Alan D.. 2023. "Google DeepMind Gemini." Life Architect (blog). May 20 [Updated Sept], 2023. <https://lifearchitect.ai/gemini/>, <https://lifearchitect.ai/gpt-5/>

This research is partially supported by the Research Council of Norway Centres of Excellence program, project number 332643, Center for Digital Narrative and project number 335129, Extending Digital Narrative.

Cite this Essay:

Johnston, David Jhave. "Embodied AI: An Extended Data Definition", Electronic Book Review, November 5, 2023, <https://doi.org/10.7273/rcss-cb21>.

Readers wishing to respond to an essay in *ebr* may send ripostes or short glosses to the journal's Managing Editor, [Will Luers](#).