Thinking of Biology

Is Biology Necessary to Advance Biology?

This manuscript is a non-peer-reviewed preprint submitted to EcoEvoRxiv.

This version is a manuscript currently under review at *BioScience*.

Although this manuscript draws on cognitive science and philosophy, its central contribution is to ecology, evolutionary biology, and conservation biology: it proposes a framework for organizing, validating, and teaching AI-assisted discovery in living systems, where knowledge claims must survive field, organismal, and cross-system constraints.

Subsequent versions of this manuscript may differ from the version posted here. If the manuscript is accepted for publication, the EcoEvoRxiv record will be updated to identify the final published article and its peer-reviewed publication DOI.

Journal under review: BioScience

Current manuscript status: First-round manuscript under review

Peer-reviewed publication DOI: Not yet available

Preprint server: EcoEvoRxiv

Date of EcoEvoRxiv submission: June 26, 2026

Thinking of Biology

Is Biology Necessary to Advance Biology?

John T. Van Stan II[1]*, Theodore Bach[2], Olivier Dangles[3,4], Robert P. Guralnick[5], Sarah E. Huebner[6], Roland Kays[7,8], Justin Kitzes[9], Robert A. Krebs[1], Benjamin J. Noren[1], Jarrad H. Van Stan[10,11,12], Michael O. Wiitala[13], Roman V. Yampolskiy[14]

[1] Biological, Geological, and Environmental Sciences, Cleveland State University, OH, USA

[2] Firelands College, Bowling Green State University, Bowling Green, OH, USA

[3] WASILAB, Centro Interdisciplinario de Ciencias de la Sostenibilidad, Pontificia Universidad Católica del Ecuador, Quito, Ecuador

[4] CEFE, IRD, University of Montpellier, CNRS, EPHE, Montpellier, France

[5] Florida Museum of Natural History, University of Florida, Gainesville, FL, USA

[6] Smithsonian Conservation Biology Institute, Front Royal, VA, USA

[7] NC Museum of Natural Sciences, Raleigh, NC, USA

[8] Forestry & Environmental Resources, North Carolina State University, Raleigh, NC, USA

[9] Biological Sciences, University of Pittsburgh, Pittsburgh, PA, USA

[10] Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA

[11] Department of Surgery, Harvard Medical School, Boston, MA, USA

[12] Massachusetts General Hospital Institute of Health Professions, Boston, MA, USA

[13] Philosophy and Political Science, Cleveland State University, Cleveland, OH, USA

[14] Speed School of Engineering, University of Louisville, Louisville, KY, USA

Author contributions: JTVS conceived the core idea at the NSF Workshop on AI, Sensors, and Ecological Infrastructure convened by JK, with initial framing developed in conversation with RPG, SEH, and RK. Case studies and broader commentary were developed collaboratively by JK, SEH, RK, RPG, RAK, and OD. The argument was then extended through discussions with JHVS on sensorimotor learning/cognition, TB on natural kinds, analogical cognition, and the epistemology of expertise, MOW on Merleau-Ponty and virtue epistemology, and RVY on AI safety, intellectology, and the limits of machine intelligence. BJN worked closely with JTVS throughout drafting, helping develop the argument and synthesize feedback from the group. All authors reviewed and revised the manuscript leading to the final version. Authorship order alphabetical.

Abstract.

AI is transforming the workflow of biological science. But what role remains for human cognition when the priority is deep theory? We propose a division of labor based on a "coverage asymmetry" between artificial and biological agents. AI excels at extending the encounterable: scaling known mechanisms and exploring complex parameter regimes. However, AI cannot (yet) perform kind formation – the minting of new variables and constraints – because it is bounded by its training ontology. Current systems cannot register that their own vocabulary is incomplete, because recognizing missing variables requires material participation in the system under study. We leverage phenomenological philosophy to argue that new kinds originate in Wild Being: the embodied friction that occurs when an organism encounters something its current categories cannot name (an Out-of-Ontology encounter). We outline a curriculum to train human scientists as "detectors of the unparameterized," ensuring that the growth of knowledge remains coupled to the growth of care. The question is not whether AI can do biology, but how we design systems to maximize discovery efficiency, producing both knowledge and knowers whose embodied engagement guides what that knowledge is for.

Key words. artificial intelligence, biological discovery, ontology, embodied cognition, out-of-distribution, epistemology.

1. The dream of BioGPT

"How many biologists – how many scientists, really – have started dreaming of a BioGPT?" At a recent NSF workshop (Kitzes et al. 2025a), a facilitator asked for a show of hands. Dozens went up. In a breakout session, the dream took clearer shape: not a Large Language Model, but a foundations model for biological theory, a Large Causal Model. A system that predicts the next causal link, not the next word.

Imagine BioGPT not only exists but delivers sustained success. As a mechanistic foundation model, it "thinks" in causal graphs of familiar variables (gene expression, canopy conductance, microbial necromass) and relations (Michaelis-Menten kinetics, Lotka-Volterra dynamics, allometric scaling), preserving competing theories with explicit uncertainty. Like AlphaFold expands what 3D protein structures are encounterable (Varadi et al. 2022), BioGPT expands what mechanistic processes are encounterable, possibly borrowing from and innovating on the diffusion models now revolutionizing protein design (Watson et al. 2023) and single-cell dynamics (Gao et al. 2025). On demand, one could run a minimal protocol: "clamp" what is securely known about a process (its nodes, units, conservation laws, and empirically constrained edges), then treat the unknown as noise, and let the model denoise toward constraint-satisfying completions, characterizing candidate mechanisms that are physically-admissible but not yet causally identified. Because this process is generative, each run yields a different completion: a physically-plausible, math-explicit, testable narrative consistent with the clamped constraints. The generated ensemble is a proposal distribution over many mechanisms (not a convergence on ground truth), for the typed, unit-consistent graphs remain undetermined by typical biological data. Each candidate is ranked by effect size and cross-system plausibility, then the best ones are subjected to lab and field testing (*i.e.*, organismal veto)—judging whether the proposed edges hold

under intervention and domain shift. Many verify. Such successes would be profound… but what kind of knowledge is being made?

2. Mechanisms and kinds: a taxonomy of biological discovery

To explore this question, consider the following taxonomy of discovery (Table 1) to distinguish between two kinds of progress, both historically familiar and valuable in biological science. The first adds testable mechanisms within the existing conceptual vocabulary, showing how known entities and activities work together to produce change (mechanistic *sensu* Craver & Darden, 2013; Machamer et al., 2000). One route recombines known mechanisms into a new chain (C1 novelty). For example, some well-known metabolic proteins regulate a gene in a different developmental pathway; these "moonlighting" proteins reveal previously unknown links (Jeffery 2018, Singh and Bhalla 2020). A second route is regime shift and conceptual transfer (C2 novelty), paralleling the logic of "transfer learning" in AI where a valid mechanistic schema is mapped onto a new domain or parameter space (Theodoris et al. 2023, Cui et al. 2025). In ecology, familiar predator-prey-producer relationships change (or flip) when ocean temperatures rise (Ashton et al. 2022, Eskuche-Keith et al. 2024) or lake trophic status changes (Carpenter et al. 2011), showing that known mechanisms in novel parameter regimes can yield qualitatively different outcomes. Because C1-C2 discoveries export existing predicates to new contexts rather than minting new ones, they are the imagined BioGPT's native territory and align with current generative-AI opportunities identified for the sciences (Karpatne et al. 2025). Even when proposed mechanistic chains originate in human "hunches" or AI "hallucinations," they remain candidates for knowledge until verification in living systems.

A second, less frequent kind of progress arises when recombination and regime exploration no longer buy explanation because the available vocabulary itself is inadequate. Here investigators stop adding links and start minting nodes in the 'knowledge graph:' the invention of new predicates or constraints ("kinds") that change what counts as an explanation (Boyd 1999, Leonelli and Ankeny 2015) (Table 1). One route is naming and operationalizing a new predicate (C3). For example, long before "trophic cascade" settled as a term, ecologists could describe wolves hunting elk and elk browsing willows, yet lacked a single predicate for why removing a top predator could restructure vegetation and alter river morphology (Paine 1980, Ripple and Beschta 2012, Ripple et al. 2016). Once named, "trophic cascade" became a portable explanatory predicate measured, and "cascade strength" supplied a measurable property for testing where and whether that predicate applies across ecological conditions (Galiana et al. 2021, Xu et al. 2023, Ripple et al. 2025). Another route is the articulation of a new cross-system constraint (C4). The early 20th century curiosity that mammalian metabolic rate scales as body mass^(3/4) (Kleiber 1932) was eventually explained by space-filling resource distribution networks (West et al. 1997) and generalized into a "metabolic scaling theory" (Brown et al. 2004, White et al. 2022)—a C4 discovery that generated portable relationships with explicit boundary conditions across terrestrial, freshwater, and marine ecosystems (Schramski et al. 2015).

Table 1 summarizes these novelty classes; Table 2 places a few recent examples from AI systems (2023-2025) on this ladder. Current biological foundation models operate comfortably at C1-C2. DECIPHAER, a C2 example, recombines known mechanisms and systematically explores neglected parameter regimes to flag regime shifts (Johnson et al. 2025). Proto-C3 examples include Evo-2 (Brixi et al. 2026) and Prov-GigaPath (Xu et al. 2024), which draft candidate kinds by learning latent features that could become predicates once named and operationalized. AI-Aristotle

(Daryakenari et al. 2024) approaches a Proto-C4 class, proposing additional terms and functional forms that may, with sufficient cross-system validation, harden into constraints. These systems can currently contribute to theory change, compressing the search over conceptual space by drafting *candidate* predicates and *candidate* functional forms, but they do not instantiate any idea end-to-end. The full pipeline from latent feature to travelling kind, or from fitted term to validated constraint, still requires a ratification process that current AI architectures do not supply: naming, operationalization, portability tests across contexts, and survival under organismal veto. This approach is the division of labor that Tables 1-2 visualize: AI now excels at extending the already-encounterable (C1-C2) and at drafting candidate kinds and constraints (proto-C3/C4); biological organisms and embedded scientists do the work of stabilizing kinds and laws.

Two cases illustrate this division. The Snapshot Safari camera-trap network demonstrates why human involvement remains essential for identifying the anomalies that precede discovery. AI scales image classification but struggles with rare species and complex behavioral contexts, often mislabeling species (*e.g.*, caracals as lions) or missing demographic details (Pardo et al. 2021). Integrating human volunteers reduced error rates from ~35% to ~9% for rare species (Huebner et al. 2024) and preserved outliers (unusual behaviors, morphs) that purely automated pipelines smoothed away (Palmer et al. 2021, Thel et al. 2021). AI manages the volume while humans preserve the anomalies that signal potential C3-C4 discoveries. A second case showed how such anomalies graduate become new predicates through embodied noticing. Around a campfire, researchers inspecting fisher (*Martes pennanti*) track-plates developed a hunch that paw-print papillary patterns were unique enough to identify individual animals. They operationalized that hunch into a protocol analogous to fingerprinting, validating that the pattern traveled reliably—transforming ‘fisher present’ into ‘individual fisher X’ (Herzog et al. 2007). The

predicate originated in a body noticing a pattern for which no variable yet existed, a classificatory handle that restructured what could be individuated and tracked.

These examples help explain why today's AI models (and our imagined BioGPT) function as powerful drafting or plausibility engines rather than autonomous discoverers of kinds. Kind formation typically demands three moves that disembodied systems do not yet supply on their own: (i) *recognizing* that the existing predicate set is inadequate; (ii) *inventing a measurement* that makes the missing variable testably real; and (iii) *separating* a portable invariant from a statistical artefact by stating boundary conditions. While models like BioGPT can recombine known mechanisms and surface candidate regularities, they currently lack the participatory history – the friction of inhabiting the system being studied – that triggers kind invention. This limitation implies that the future of discovery requires more than scaling computation; it points toward a more complete cognitive architecture that is embodied, embedded, enactive, and extended.

3. How bodies, tools, and places make biological knowledge

To understand why kind formation resists automation, consider how biological knowledge emerges. A soil ecologist disturbs leaf litter in a familiar forest plot and notices an unusual earthy-sweet odor. This sensory anomaly prompts questions, new sampling protocols, and reveals a phenomenon that eventually gets formalized through instruments (like chemical analyses, microbial assays). What began as embodied pattern recognition becomes a testable hypothesis about microbial metabolism under specific environmental conditions. This example illustrates 4E cognition (Newen et al. 2018, Carney 2020, Alexander 2025), a cognitive science framework that recognizes thinking is not a detached, internal process ("in the head"), but emerges from a coupled system that is: (i) Embodied in sensory capacities (the ecologist's olfactory detection); (ii)

Embedded in specific environments (this forest, moisture, and temperature regime); (iii) Enacted through intervention and feedback (sampling and monitoring); and (iv) Extended by tools (chemical analyzers, statistical frameworks).

AI systems like AlphaFold exemplify "extended" cognition, a computational tool that widens what a research community can encounter without altering the field's conceptual grammar. AlphaFold predicts three-dimensional protein structures across entire proteomes (Varadi et al. 2022), making an enormous region of mechanism-space newly encounterable: a spectacular C1-C2 expansion. What it does not supply is the capacity to register that the conceptual vocabulary guiding its predictions is itself incomplete—the recognition that triggers kind formation (C3-C4). This missing capacity is called here, 'embodied friction,' or the material coupling between an inquiring organism and its study system that generates expectations the organism did not explicitly encode, and that forces ontological repair when those expectations are violated.

Terrestrial bioacoustics illustrates how embodied friction operates alongside AI extension. Researchers position microphones using knowledge of vegetation density, wind, and animal corridors – their bodies and field experience "embedding" inquiry in real constraints. AI detection models then "extend" perception across thousands of recording hours and frequency ranges beyond what human hearing can parse (Kitzes et al. 2025b). But it is the researcher who recognizes when a dawn chorus "sounds wrong" in context-dependent ways that algorithms trained on statistical baselines often miss or misclassify. That recognition (an embodied expectation violated by what the ear encounters) is the friction that flags where existing predicates may be inadequate and new ones may be needed. Without that embedded and embodied oversight, error patterns shift silently with background noise and microphone placement, and the conceptual vocabulary remains unchallenged. Seen this way, 4E cognition becomes more than a philosophical add-on but a

working description of how biology continues to make sense of itself. Tools and algorithms extend the reach of attention; field sites and model organisms embed inquiry in real constraints; experiment and manipulation enact understanding through feedback; and skilled minds embody the sensing and judgment that transform surprise and conjecture into knowledge. Together they define a cognitive ecology in which human and artificial agents cooperate most fruitfully—AI extends the encounterable (scale and speed), while embodied minds determine what is worth pursuing.

Given this objective, why not just embody the AI? Imagine BioGPT-bot; a robot equipped with advanced sensors to navigate terrain and sample physical and biogeochemical parameters adaptively (*sensu* Henrys et al. 2024). It even "learns" new protocols. While technically the bot occupies space, navigates terrain, and suffers humidity, cold, and power limits, it lacks something fundamental for discovery: sensibility (the metabolic vulnerability of a living system). A biologist's body is a living system studying living systems, entrained to the same light cycles that structure organism behavior, stressed by the same temperatures that constrain plant physiology, and responsive to the same humidity gradients that shape fungal growth. Human thermoregulation, fatigue, smell, heading, and proprioception are generic measurement devices only secondarily; primarily, they are biological responses shaped by the very evolutionary-ecological pressures that shape the organisms under study. That shared history creates pre-attentive saliences (what feels "off" or "alive") that sensors cannot (perhaps yet) replicate. BioGPT-bot records data; the biologist's body notices what had not yet been known. That gap between extending known protocols and recognizing that the protocols themselves are inadequate marks the frontier where kind formation begins.

4. Wild Being: When life studies life

Before arguing that embodied encounters supply something AI currently does not, let's steelman the alternative. Imagine a sophisticated AI-sensor network on an ocean buoy that detects a shift in pH and dissolved phosphorus, relays it to an agentic system that assimilates decades of time-series data, and issues a probabilistic forecast of a plankton bloom, updating itself when new sensor data contradicts the forecast. Knowledge-guided architectures now incorporate domain ontologies and physical constraints into model training and inference, uncertainty quantification methods flag when models operate outside their reliable range, and complementary AI paradigms (from symbolic reasoning to reinforcement learning) are being integrated to improve consistency and decision-making (*cf.* Karpatne et al. 2025). This is no imaginary system. Its component parts and their integration represent genuine advances in extending the encounterable and, increasingly, in drafting candidates for new kinds (in addition to Table 2). Additionally, human embodied intuition can be unreliable. Anthropocentric assumptions can misdirect cross-species inference (Brebner et al. 2024, Barton and Barrett 2025), clinician "hunches" underperform statistical models in well-defined prediction tasks (Dawes et al. 1989, Grove et al. 2000), and ethicists have misused the "wisdom of repugnance" to impede legitimate inquiry, particularly in biotechnological research (Kass 1997; cf. Nussbaum 2004). Fundamentally, shared material history is a filter as much as a channel, systematically foregrounding perturbations to which the observer's own biology is attuned (Wilson 2002, Van Dyck 2012). Even so, discovery is bottlenecked not only by correctness but by what gets noticed as worth formalizing. Wild Being, as developed below, is not a truth-guarantee or inference engine but a salience channel; when it tracks something real, it can initiate naming, instrument design, and boundary-condition mapping in ways current AI does not

yet originate on its own. When it misfires, the organismal veto (Box 2) and AI-based checks help discipline false positives and parochialism.

With this concession in place, consider three scenes. A marine biologist slips into a familiar reef and notices the water surface feels oddly 'tacky' hours before a plankton bloom. An ornithologist registers a missing voice in the dawn chorus before surveys document territory abandonment. A botanist touches leaves and recognizes incipient drought stress days before wilting becomes visible to cameras or satellites. None of these are measurements or parts of established protocols. They are pre-predicative recognitions enabled by biological similarity. The marine biologist's tactile and olfactory systems did not evolve to detect precursors of dinoflagellate blooms, but they were tuned in aquatic environments where lipid chemistry, surface tension, and dissolved compounds structure sensory experience. The ornithologist's auditory processing evolved in acoustic landscapes that included avian vocalizations. The botanist's mechanoreceptors respond to turgor changes because plant and animal cells regulate water through overlapping biophysical principles. In each case, shared evolutionary and material history makes some patterns salient to a living observer before they are named or measured.

Philosopher Maurice Merleau-Ponty – who engaged with the science of his time, including psychology, developmental biology, and physics (Seglard 2003, Carman and Hansen 2004) – called this phenomenon *l'être sauvage*: Wild Being, or flesh (Merleau-Ponty 1969, 1992). He uses the term "Wild Being" to name the shared material continuity in which living systems articulate and reveal themselves to one another.[1] When the biologist feels the reef is "off," they are not inferring a conclusion; their biological system is registering a material shift in the local biological

[1] See Merleau-Ponty, *The Visible and the Invisible* (1969), "Working Notes" (May 1960), esp. p. 248, where he characterizes "l'être sauvage" as a pre-predicative order shared by perceiver and world ("l'être sauvage n'est pas opacité, mais dimensionnalité"), mechanistically grounding why certain patterns are perceptible to living observers before they are measured.

system—a "mutual experience" possible because observer and observed are "fundamentally homogeneous"[2] (*i.e.*, formed by the same evolutionary processes). Some brief philosophical clarification is needed. Wild Being is not the "preconceptual given" that philosophers like Immanuel Kant and Wilfrid Sellars rejected (Triplett 2023). Kant established that "intuitions without concepts are blind" (Kant 1787). Sellars (1956) extended the argument by showing that nothing counts as knowledge unless it is already situated in inferential relations among concepts—the "logical space of reasons." We do not dispute this. Kant and Merleau-Ponty agree that conceptual activity is already operative in sensory experience. Merleau-Ponty parts from Kant in arguing that beneath both stabilized concepts and sense qualities lies a dimension of embodied sense-making from which both emerge, an embodied intelligibility before it is tamed into concepts (what Merleau-Ponty elsewhere calls *sens sauvage*). Wild Being names this pre-predicative field—not separate from conceptual mediation, but a source of sense that prompts the generative reorganization of concepts when the world refuses the available categories. Our argument requires only this minimal commitment: that there exists a relational, embodied register through which living systems sense perturbations in one another before stabilized instruments and predicates catch up.

This outcome also differs from "common sense biology" or naive intuition (*sensu* Wolpert 1994). Mature science may appear "unnatural" in routinely overturning folk expectations, but even the most counter-intuitive theories are disciplined by friction with living systems. Wild Being, as used here, names the pre-predicative channel where such friction first appears—e.g., the chemistry of rotting cactus (Fogleman and Danielson 2001) cohering as a recognizable syndrome of tactile, olfactory, and visual cues. These recognitions mark where current concepts are being challenged

[2] See Merleau-Ponty, The Visible and the Invisible (1969), "Interrogation and Intuition," p. 114.

by the external world. In this sense, Wild Being is the precondition to the "unnatural" character of science. There is no route to genuinely revisional theory that does not run through situated encounters where the world 'refuses' our expectations. Operationally, Wild Being is the spark that opens a worklist – flags the anomaly, sketches a feature, and engineers a protocol – driving C3 and C4 novelty. For C3 (kind formation), the biologist notices an anomaly in Wild Being (the tacky water surface), names it through conceptual invention, then measures change through instrumental formalization so that the candidate predicate gets a protocol others can run. For C4 (new constraints), the scientific task is to hunt for the breaking point, *i.e.,* a series of structured surprises in places where a trusted relation unexpectedly fails across different sites or conditions. Once a frontier of failure is traced, observations of systematic breakdown can be formalized into boundary conditions, compressing the insight into a portable law with explicit thresholds (often with hysteresis).

Why can't current AI, scaffolded with physics, biology, unit consistency, etc., detect such violations? Because there are two kinds of rules at work. A model's scaffold currently encodes explicit, formal constraints (thermodynamics, stoichiometry, mass balance) (Karniadakis et al. 2021, Faure et al. 2023). A biologist's body develops implicit expectations, a high-dimensional sense of how this reef or valley or canopy should behave. An AI's "surprise" is a statistical anomaly against its model, a deviation within the parametrized variable space on which it was trained. A biologist's surprise is a perturbation registered by a sensory baseline built from material co-participation with the system, a baseline that extends, noisily and unreliably, beyond any current parametrization. Both are pattern-matching engines, and within a well-specified ontology the AI is often superior at this point. But the anomalies that seed C3/C4 discovery are precisely those for which no variable yet exists. There, detection depends on whether the agent's baseline has

coverage in the missing dimension, and a baseline built from shared material constraints sometimes does, where one built from explicit parameters, by construction, cannot (today).

This distinction leads to a central hypothesis: that C3-C4 discovery requires constraint-sharing (being subject to the same thermodynamic, temporal, and ecological pressures as the study system). This concept is a testable claim: do conceptual breakthroughs correlate with the researcher's material overlap with the study system? If so, disembodied AI faces a principled barrier at the C3/C4 frontier; not a general barrier to discovery, but a barrier to types of kind-formation that depend on material overlap with the system under study. If not, the barrier is merely practical, reflecting the current absence of constraint-sharing in generative models. When extrapolating beyond trained data, models may fabricate plausible outputs within their parameterized space rather than signal that the space itself is incomplete (Huang et al. 2025). AlphaFold 3's occasional stereochemical errors – chirality flips, atomic clashes, "ordered hallucinations" in intrinsically disordered regions – are instructive here, showing that the model can optimize confidently where its ontology has no coverage, producing physically-admissible but chemically-fictitious structure (Abramson et al. 2024, Magana Gomez and Kovalevskiy 2024). A biologist handling the same protein might not predict the fold, but their biochemical intuition (trained on years of gel behavior, binding kinetics, and crystallization failures) could register that something about the output "doesn't sit right," because the human baseline extends into dimensions the model does not parameterize. A scientist, by contrast with the model, must triage anomalies, deciding which deviations are noise, which are artifact, and which may indicate that the existing frame is incomplete—knowing that this judgment shapes the kind of scientist one becomes. This is a form of embodied virtue epistemology, the skillful anomaly triage cultivated

through practice and accountability that is not reducible to noise reduction. Wild Being thus opens the worklist, then today's AI can sweep the charted territory.

One might object that these limitations are engineering rather than epistemic, solvable by fine-tuning on physics, chain-of-thought scaffolds, or coupling methods to simulated environments. The literature does document such mitigations and reports incremental gains (Song et al. 2026). More ambitiously, orchestration frameworks pit multiple models against one another and against continuous data streams, creating consistency loops of genuine power (Gao et al. 2024, Ghafarollahi and Buehler 2025). But consistency is not coverage. Adding models improves cross-validation within a shared ontology, not sensitivity to dimensions beyond it. An orchestrated BioGPT-bot could sample more data yet still fail to notice that the sampling regime is missing a necessary variable. The 3D-reasoning failures catalogued by Song et al. (2026) make this point concrete: even physically-situated LLM-driven agents – robots interacting with environments and receiving sensorimotor feedback – still produce systematically impossible action plans (Song et al., 2026), a deficit traced to the absence of a "robust internal world model" (*sensu* Dao and Vu 2025, Wu et al. 2025). Current systems are increasingly adept at out-of-distribution (OOD) detection, flagging data that are statistically improbable given known parameters. C3/C4 discovery, however. addresses a different problem: an out-of-ontology (OOO) encounter, where the model lacks the dimension required to represent what it encounters. To a system parameterized by pH and temperature, the "tackiness" of the water is not an outlier but a null value (there is no axis on which to register a deviation). Engineering can mitigate for OOD by expanding training distributions; OOO requires an agent whose baseline extends into the missing dimension. Wild Being, as we have defined it, names the capacity to register this friction and be perturbed by a variable that has yet to receive a name.

For now, the optimal human-AI system for biology is one that exploits this coverage asymmetry as a design resource. AI's computational power extends the already-encounterable by exploring known variable space at unprecedented scale, while embodied cognition supplies anomaly detection in the unparameterized margins where new variables originate. Without this embodied anchor, an "inaction vortex" of place-blind data may occur where disembodied monitoring substitutes for intervention (Devictor and Bensaude-Vincent 2016). Beyond method, Wild Being generates ethical commitments that autonomous systems cannot replicate today. When researchers spend months in an ecosystem, they (ideally) develop attachments to places and stakes in the system's futures—a root of scientific accountability that helps ensure knowledge production serves more than efficiency metrics and that questions of value are not delegated to optimization alone.

5. Wild attachments: The ethical tether and existential pull of embodied biology

Removing bodies from places risks more than missed C3-C4 discovery; it invites place-blind inference. Models trained "there" could fail "here" without notice because no agent has "skin in the game" (*i.e.*, a material stake in the outcome) to ask e.g., *which questions matter?, which risks are acceptable?*, and *which futures matter?*. This omission invites circular validation, where disembodied systems confirm their own assumptions, optimizing for metrics that may not align with ecological or human flourishing. It also risks foreclosed discovery, where the anomalies that have historically forced conceptual revision go unregistered. These limits are bigger than epistemic failures, they are ethical ones because they sever the feedback loop between knowledge and care. Even if future iterations of BioGPT-bot achieve technical embodiment (sophisticated sensors,

adaptive morphology, enactive algorithms, embedded constraints) an existential question remains: What is biological science for?

Avoiding the abstract, one working answer has emerged, almost accidentally, from a platform that embodies the design principle that Section 4 advocates. In contrast to the disembodied "view from nowhere" of global biodiversity databases, iNaturalist (*www.inaturalist.com*) cultivates a form of wild attachment through the productive friction of machine-human interaction. While technically a data infrastructure (Leonelli 2016), iNaturalist's practical effects are often the opposite of epistemic distancing: re-embedding users in multispecies encounters by mobilizing AI to activate field experience (Mason et al. 2025). AI mis-identification errors – especially for species with few database records (Mesaglio et al. 2025) – can be productive. They trigger debates among taxonomists, naturalists, and casual observers, prompting return visits, additional photo uploads, and documentation of context that escapes pattern-recognition systems (*e.g.*, smells, habitats, associated practices, local names). Each misclassification thus catalyzes a co-production loop in which the encounter between observer and organism is not merely archived but continually re-enacted through comment threads, anecdotes, accounts of local uses, and debates over diagnostic traits (Campbell et al. 2023). The system's indeterminacies become conditions for wild attachment by inducing a return to the field, offering a counterpoint to the "inaction vortex" produced when remote monitoring substitutes for place-based engagement (Devictor and Bensaude-Vincent 2016). The design principle here is not to keep AI peripheral; rather, AI is most virtuous (most conducive to good science and good scientists) when it intensifies return-to-field loops and community deliberation. This outcome is a different ethical architecture than AI-as-autonomous-discoverer and suggests that how we deploy AI matters as much as what AI can do.

This approach again reinforces the realist principle that knowledge production is never purely instrumental. When humans study living systems, they produce situated interactions, attachments, and practices that sustain them. Here, the BioGPT-bot thought experiment deserves a second look, to identify precisely what is absent functionally. BioGPT-bot optimizes task completion where "values" are parameters encoded at deployment. Whether AI systems could ever develop something functionally analogous to caring, vulnerability, or metabolic stake is an open question and, in today's material conditions, these absences have specific epistemic consequences that connect to the coverage argument above. Caring is what makes a scientist persist with a recalcitrant anomaly long enough to push it through the C3/C4 pipeline, perhaps returning to the reef season after season until the "tacky" feeling becomes a named chemical signature. Vulnerability is what gives the anomaly a felt urgency coupled with a potential risk in the same environment suggesting precisely why discomfort carries information. These are functional roles in the discovery process, not sentimental extras. A system that lacks care is incomplete as a discovery architecture, which is why a human-AI partnership, rather than AI autonomy, remains the appropriate design frame. Biology should remain a practice that (at its best) reorients the knower.

6. Curricula for combining two kinds of knowledge.

The difference between how a practicing biologist would train a BioGPT-bot and how they ought to train a human student distills the argument of this essay. The training curriculum for BioGPT is reassuringly orthodox. First, pretrain on the canon: formalize mechanisms from textbooks and ODE (ordinary differential equation) models into typed, unit-consistent graphs scaffolded by constraints like stoichiometry. Then, mask and diffuse. Whether via BERT-style

masking or diffusion-based denoising, the objective is the same: train the model to predict missing links consistent with the provided data. Finally, evaluate based on the recovery of known pathways (optimizing for C1-C2 performance). Semantic reasoning modules can extend this approach by drawing inferences over structured ontologies (Aleksander et al. 2026), and meta-learning architectures can flag latent structural mismatches within the model's own representations (Shi et al. 2025)—genuine advances that push toward proto-C3 capabilities (Section 2, Table 2). But the ontology over which these systems reason, and the distribution against which they detect mismatches, are still defined by the training regime. The model can discover that parameters are inconsistent but not discover an incomplete vocabulary, because incompleteness, by definition, lies outside the parameterized space against which the loss is computed. Whether future architectures could develop baselines with coverage beyond their training ontology (achieving something functionally analogous to material co-participation) is an open question.

A human student, by contrast, is someone to train to move work from C2 into C3/C4, often through collective sensemaking and sustained engagement with others (a capacity that, for now, requires a different architecture entirely). They must (i) cultivate orthodoxy (to speak the grammar), (ii) develop baseline depth (to detect when the grammar fails), and (iii) sustain engagement (to persist through the work of stabilization). We cover each below in more detail.

(i) *Train the canon and its limits.* Students need deep familiarity with mechanisms that are taught as a record of past kind-formation with visible blind spots. Instructors should use the taxonomy of discovery to interrogate classic papers. *What was <u>not</u> in the literature that required these authors to mint a new term?* This style teaches students that "orthodoxy" is a frozen history of past kind-formations, and that every current vocabulary has similar gaps waiting to be exposed.

(ii) *Build deep baselines through staged encounters.* The coverage argument from Section 4 has a pedagogical corollary: anomaly detection in unparameterized dimensions requires a sensory baseline built through sustained material co-participation within a system. Field courses and long-term lab projects should require "anomaly logs"—running records of observations (smells, textures, or behaviors) that resisted easy classification under existing variables. Seminars can then push these through the C3 pipeline. This pipeline could begin with (a) sketch a predicate for the anomaly, (b) imagine how it might travel, (c) identify where it would cease to apply. Students should be sent on C4-hunts (to find where taken-for-granted relationships, *e.g.*, size-temperature scaling) break down. The goal is to build the kind of baseline that extends, however imperfectly, beyond current parameterizations, so that when a violation occurs in an unnamed dimension, the student has enough coverage to register an anomoly.

(iii) *Sustain engagement to enable stabilization.* Kind formation is slow. A predicate must be coined, operationalized, tested across contexts, and defended under the organismal veto. Such a pipeline unfolds over years (not semesters). Long-residency projects build the baseline depth that makes anomaly detection possible, and they build the commitment to see a candidate kind through the full stabilization process. Risks remain; for example, deep familiarity with one system can produce parochialism, where a locally valid pattern is mistaken for a traveling kind. Fortunately, this limitation in a human is precisely where AI serves as a corrective—testing whether a predicate coined in one valley holds across the variable space of a thousand others. The partnership runs both ways, where embodied engagement generates the candidate and computational breadth disciplines the scope.

While the above may appear to wall-off BioGPT, in practice the integration of BioGPT is straightforward. Teach students to see AI as an Extended tool in a 4E architecture. Teachers can

use AI to scan parameter spaces while assigning students the job of designing ground-truth experiments to adjudicate the output. Side-by-side analyses, where students compare AI predictions with their own anomaly logs, become teachable moments about the C3/C4 frontier and about the complementary coverage of the two systems. Still, this pedagogical vision faces challenges regarding how to assess baseline depth and anomaly sensitivity in an academy that rewards quantifiable outputs. The prolonged noticing that precedes publication and the failed experiments that map constraint boundaries are rarely legible in current evaluation frameworks. Yet, these omissions are precisely what distinguish discoverers from data processors. An educational program that treats students as future BioGPTs – maximizing canonical data consumption while minimizing time in meadows and wet labs – will produce skilled link-adders, but they are unlikely to produce the creators who notice missing data on a known axis and recognize when the axis itself is absent. Do not shelter students from AI, rather, train them to inhabit a different role in the cognitive ecology, as the biological agents whose baselines extend into the unparameterized margins where new variables originate.

7. Conclusions: Designing for productive asymmetry

The future of biological discovery depends on how the cognitive ecology between human and artificial agents is designed. The two systems currently have different coverage profiles: AI excels within parameterized variable space, while embodied inquiry extends, imperfectly, into unparameterized dimensions where new variables originate. Good design exploits this asymmetry. The division of labor propose here is grounded in 4E cognition and is explicitly revisable as capabilities evolve:

- **Keep the Extended strong**: Use AI to generate coherent candidate mechanisms (C1) and sweep parameter regimes (C2) at scales no human team could process. AI's strength is coverage within the current ontology. Leverage it fully.
- **Protect the Embedded**: Ensure that models trained in one context are governed by researchers embedded in the contexts where they are applied. Domain shift (where a model trained "there" fails "here") is more than a statistical problem; it is an epistemic one. Local validity depends on knowledge that resists compression into portable parameters. Embedded researchers serve as governors of this local validity.
- **Make it Enactive:** Treat AI predictions as probes designed to stress-test the ontology itself. The standard cycle (predict, test, revise parameters) improves C1-C2 performance. The additional move advocated here is deliberate C4-hunting. Deploy predictions across contexts specifically to find where trusted relationships break, then use those failures to map constraint boundaries. Iteration should aim not only at better predictions within the current variable space but also expose where that space has holes.
- **Fund the Embodied**: Treat sustained material engagement with study systems (in field sites, in long-running labs, at instrument benches) as essential infrastructure for discovery. This proposal does not mean to claim that fieldwork is privileged over benchwork, but to encourage baseline depth, built through prolonged co-participation with a system, which is the most reliable way for building the sensory coverage that detects anomalies in unparameterized dimensions. Where that co-participation occurs (a reef, a sequencing facility, a greenhouse) matters less than providing an opportunity at a depth and duration required for baseline formation. In practice, funding agencies and institutions should prioritize long-term, place-based projects that cultivate embodied expertise over short-cycle deliverables, curricula should

train students to take responsibility for study systems (not just extract data from them), and evaluation criteria should reward the baseline depth and anomaly sensitivity that precede publishable discovery (cf. Dangles 2023)

This division of labor carries ethical weight alongside epistemic function. Discovery generates knowledge and commitment. Researchers whose sustained co-participation builds baseline depth simultaneously develop the attachments that determine which questions feel urgent and which risks are acceptable. Whether future architectures could develop out-of-ontology sensitivity is an open empirical question; nothing in our framework defines Wild Being as exclusively biological and if such systems emerge, the division of labor would need to be redrawn. But this argument about organizing epistemic labor in biology is calibrated to the present asymmetry. For as long as extending the encounterable and inventing what is encounterable require different coverage profiles (one computational, one biological), this division of labor holds. That coordination will be slower, costlier, and messier than automation alone for now. It is also, to date, the only path that has generated discoverers who loved what they found enough to fight for its survival.

Acknowledgements. NSF award #2213623 supported JTVS and BN. NSF #2528520 supported JK's workshop. Thanks to the University of Pittsburgh Hub for AI and Data Science Leadership for hosting.

Table 1. A taxonomy of discovery in AI-assisted biological science. This includes novelty classes, who (typically) makes them, what counts as evidence, and each class's epistemic function.

Novelty class	Description	Typical AI contribution	Typical embodied contribution	What counts as evidence?	Epistemic function
C0: Retrieval *Ex:* Literature retrieval	Restates a known mechanism from existing literature.	Aligns prior causal links; checks consistency across a knowledge graph.	Curates the relevance of known facts to a specific, local context.	Matches literature; reproduces an expected outcome *in situ*.	Keeps the 'obvious' obvious; confirms shared ground.
C1: Mechanism addition *Ex:* Moonlighting proteins	Builds a new causal chain from previously unconnected but known components.	Generates coherent, testable hypotheses by combining existing mechanisms.	Designs feasible assays; performs skilled sampling in the right place and time.	Verified causal effect and the operative mechanism *in vivo* or *in situ*.	Expands the encounterable into the actualizable.
C2: New regime *Ex*: Prey-predator relationship in hotter system	Shows how known mechanisms yield a qualitatively new outcome in a novel parameter range.	Systematically explores parameter space; flags potential regime shifts and boundary conditions.	Establishes the new regime in a living system; tracks historical contingency and path-dependence.	A reproducible regime shift with clear thresholds or hysteresis.	Deepens theory by revealing the constraints on existing models.
C3: Kind formation *Ex*: Trophic cascade	Names a new, measurable variable or predicate that becomes explanatory.	Scales the new "kind" once named and operationalized; integrates it into pipelines.	Coins the predicate from a surprising encounter; invents the measurement or instrument.	The new "kind" travels across sites/systems and becomes a stable explanatory concept.	Changes the grammar of biological explanation.
C4: New law or constraint *Ex*: Metabolic scaling	Proposes a cross-system invariant or scaling law that reorganizes entire models.	Derives the candidate form; suggests tests across different systems/modalities.	Designs and executes the multi-site trials; probes for failure and exceptions that clarify the law's boundaries.	Holds true across taxa or biomes with precisely defined boundary conditions.	Rewires theory by adding a durable, high-level structure.

Table 2. The autonomy horizon: Distinguishing AI optimization from kind formation. Current AI systems excel at exploring the map (C1-C2) but function as "engines" for new concepts (Proto-C3, Proto-C4), requiring embodied ratification to be knowledge.

Novelty Class	Representative Models	Epistemic Status	The Missing "Wild" Link
C2: Regime Shift	**DECIPHAER** (Johnson et al. 2025). Identifies a dose-dependent switch where bacteria transition from cell-wall stress (ATP↑) to respiration inhibition (ATP↓).	**Parameter space explorer** (high autonomy). AI systematically explores the state space of known variables to identify qualitative transitions and regime boundaries. It discovers new coordinates but operates within existing dimensions.	**Verify the regime.** Requires embodied validation of regime in living systems (not just model artifact) and mapping where transitions occur under real biological constraints (pH, temp, etc.).
Proto-C3: Candidate predicates/scores that might become explanatory kinds	**Evo-2** (Brixi et al. 2026) **Prov-GigaPath** (Xu et al. 2024) **(**Genomic/Pathology FMs). Surfaces "latent features" or "patho-contextual archetypes" that correlate with outcomes but lack semantic labels.	**Drafting engine** (candidate kinds). Models identify mathematical clusters. However, a cluster is not a biological "kind" until it is: (1) semantically named, (2) operationalized via portable protocols, and (3) validated to travel across contexts.	**Stabilize the kind.** Requires embodied recognition that a cluster corresponds to a biological reality; e.g., transforming "Latent Feature 247" into "Cryptic Stress Response" through sensory-ecological encounter.
Proto-C4: New Constraint Candidates	**AI-Aristotle** (Daryakenari et al. 2024) (Symbolic Regression). Discovers "hidden physics" terms to improve ODE fit for pharmaco-kinetics, suggesting missing regulatory constraints.	**Proposal engine** (candidate laws). AI surfaces candidate functional forms (e.g., "Parameter X scales with Y"). These begin as statistical regularities or engineering rules, not yet established as biological laws.	**Validate the invariant.** Requires the Organismal Veto: testing the proposed rule across diverse taxa and biomes to see where life respects v. violates it, thereby establishing the law's true boundary conditions.

BOX 1: The "Turing Test" for AI discovery

What Would AI Achieving C3/C4 Discovery Look Like?

We argue that current AI acts as a "drafting engine" for biological discovery. To qualify as an autonomous discoverer of kinds (C3) or constraints (C4), rather than just an optimizer of existing ones, an AI (not just its human handlers) must:

1. Coin a Predicate: Invent a variable not present in its training ontology (e.g., "trophic cascade strength") and provide a first operationalization (a measurement protocol) that produces discriminating data *in vivo*.

2. Demonstrate Portability: Show that this new predicate travels across ≥2 distinct biomes or systems (e.g., from wolves to sea otters) with stated boundary conditions, correctly identifying where the concept applies and where it fails.

3. Articulate a Cross-System Constraint: Derive a governing rule (C4) whose parameters are estimated independently from new data, and which systematically reorganizes existing models (e.g., improving prediction across taxa by adding a metabolic term).

4. Survive the Organismal Veto: The predicted effect must replicate in a living system where domain shift, history, and noise are not controlled away.

Until a system achieves these four steps unaided, we treat AI as maximally valuable for extending the encounterable (C1-C2) and as a force multiplier for humans stabilizing the new (C3-C4).

BOX 2. The organismal veto: How life rejects bad hypotheses

The "Organismal Veto" is the process where a theoretical prediction is rejected by the living system. This veto operates differently across the taxonomy of discovery, distinguishing genuine biological signals from model artifacts.

C1 (mechanism): Veto of context.

- Ex. AI prediction: Protein *X* binds to promoter *Y* based on high motif similarity.
- The veto: The interaction works *in vitro* but fails in living cells due to chromatin accessibility or compartmentalization.
- Epistemic gain: Reveals that mechanisms require structure in context; structure alone is insufficient.

C2 (Regime Shift): Veto of resilience.

- Ex. AI prediction: Ecosystem will flip from kelp forest to urchin barren at 2 °C warming
- The veto: The system absorbs 3 °C through phenotypic plasticity and compensatory predation, maintaining its state.
- Epistemic gain: Reveals hidden robustness and compensatory dynamics operating outside the model's parameter space.

C3 (Kinds): Veto of coherence.

- Ex. AI prediction: Cluster 247 in a single-cell foundation model represents a novel cell type.
- The veto: The pattern dissolves when measured across different batches, tissues, or organisms—it is confirmed to be a statistical artifact, not a biological reality.
- Epistemic gain: Forces the distinction between a statistical cluster and a traveling biological kind.

C4 (Constraints): Veto of exception.

- Ex. AI prediction: A universal scaling law applied to all animals.
- The veto: Flatworms and tardigrades systematically violate the exponent due to surface-area-dominated geometries and cryptobiotic metabolic modes.
- Epistemic gain: These violations are not noise; they map the law's boundary conditions, converting a correlation into a structural constraint.

References.

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung C-C, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Žídek A, Bapst V, Kohli P, Jaderberg M, Hassabis D, Jumper JM. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630: 493–500.

Aleksander SA, Balhoff JP, Carbon S, Cherry JM, Ebert D, Feuermann M, Gaudet P, Harris NL, Hill DP, Kalita P, Lee R, Mi H, Moxon S, Mungall CJ, Muruganujan A, Mushayahama T, Sternberg PW, Thomas PD, Van Auken K, Wong ED, Wood V, Ramsey J, Siegele DA, Chisholm RL, Dodson R, Fey P, Aspromonte MC, Nugnes MV, Naser XAC, Tosatto SCE, Giglio M, Nadendla S, Antonazzo G, Attrill H, Brown NH, dos Santos G, Marygold S, Röper K, Strelets V, Tabone CJ, Thurmond J, Zhou P, Zaru R, Lovering RC, Logie C, Chen D, Naba A, Christie K, Corbani L, Ni L, Sitnikov D, Smith C, Seager J, Cooper L, Elser J, Jaiswal P, Gupta P, Naithani S, Carme P, Rutherford K, De Pons JL, Dwinell MR, Hayman GT, Kaldunski ML, Kwitek AE, Laulederkind SJF, Tutaj MA, Vedi M, Wang S-J, D'Eustachio P, Aimo L, Axelsen K, Bridge A, Hyka-Nouspikel N, Morgat A, Goldbold G, Engel SR, Miyasato SR, Nash RS, Sherlock G, Weng S, Bakker E, Berardini TZ, Reiser L, Auchincloss A, Argoud-Puy G, Blatter M-C, Boutet E, Breuza L, Casals-Casas C, Coudert E, Estreicher A, Famiglietti ML, Gos A, Gruaz-Gumowski N, Hulo C, Jungo F, Mercier P Le, Lieberherr D, Masson P, Pedruzzi I, Pourcel L, Poux S, Rivoire C, Sundaram S, Bateman A, Adesina A, Bowler-Barnett E, Carpentier D, Denny P, Ignatchenko A, Ishtiaq R, Lock A, Lussi Y, Magrane M, Martin MJ, Orchard S, Raposo P, Speretta E, Tyagi N, Urakova N, Warner K, Yu CW-H, Chan J, Diamantakis S, Quinton-Tulloch M, Raciti D, Fisher M, James-Zorn C, Ponferrada V, Zorn A, Howe D, Ramachandran S, Ruzicka L, Westerfield M. 2026. The Gene Ontology knowledgebase in 2026. *Nucleic Acids Research* 54: 1779-1792.

Alexander C. 2025. What is 4E cognitive science? *Phenomenology and the Cognitive Sciences* 10.1007/s11097-025-10055-w.

Ashton G V., Freestone AL, Duffy JE, Torchin ME, Sewall BJ, Tracy B, Albano M, Altieri AH, Altvater L, Bastida-Zavala R, Bortolus A, Brante A, Bravo V, Brown N, Buschmann AH, Buskey E, Barrera RC, Cheng B, Collin R, Coutinho R, De Gracia L, Dias GM, DiBacco C, Flores AAV, Haddad MA, Hoffman Z, Erquiaga BI, Janiak D, Campeán AJ, Keith I, Leclerc JC, Lecompte-Pérez OP, Longo GO, Matthews-Cascon H, McKenzie CH, Miller J, Munizaga M, Naval-Xavier LPD, Navarrete SA, Otálora C, Palomino-Alvarez LA, Palomo MG, Patrick C, Pegau C, Pereda S V., Rocha RM, Rumbold C, Sánchez C, Sanjuan-Muñoz

A, Schlöder C, Schwindt E, Seemann J, Shanks A, Simoes N, Skinner L, Suárez-Mozo NY, Thiel M, Valdivia N, Velez-Zuazo X, Vieira EA, Vildoso B, Wehrtmann IS, Whalen M, Wilbur L, Ruiz GM. 2022. Predator control of marine communities increases with temperature across 115 degrees of latitude. *Science* 376: 1215-1219.

Barton R, Barrett L. 2025. Embodied cognitive evolution and the limits of convergence. *Philosophical Transactions of the Royal Society B: Biological Sciences* 380: 20240255.

Boyd R. 1999. Homeostasis, Species, and Higher Taxa. Ch. 6 in *Species: New Interdisciplinary Essays* (Ed. Wilson R.A.): MIT Press.

Brebner JS, Loconsole M, Hanley D, Vasas V. 2024. Through an animal's eye: The implications of diverse sensory systems in scientific experimentation. *Proceedings of the Royal Society B: Biological Sciences* 291: 20240022.

Brixi G, Durrant MG, Ku J, Poli M, Brockman G, Chang D, Gonzalez GA, King SH, Li DB, Merchant AT, Naghipourfar M, Nguyen E, Ricci-Tam C, Romero DW, Sun G, Taghibakshi A, Vorontsov A, Yang B, Deng M, Gorton L, Nguyen N, Wang NK, Adams E, Baccus SA, Dillmann S, Ermon S, Guo D, Ilango R, Janik K, Lu AX, Mehta R, Mofrad MRK, Ng MY, Pannu J, Ré C, Schmok JC, John J St., Sullivan J, Zhu K, Zynda G, Balsam D, Collison P, Costa AB, Hernandez-Boussard T, Ho E, Liu M-Y, McGrath T, Powell K, Burke DP, Goodarzi H, Hsu PD, Hie BL. 2026. Genome modeling and design across all domains of life with Evo 2. *Nature* 10.1038/s41586-026-10176-5.

Brown JH, Gillooly JF, Allen AP, Savage VM, West GB. 2004. Toward a metabolic theory of ecology. *Ecology* 85(7): 1771-1789.

Campbell CJ, Barve V, Belitz MW, Doby JR, White E, Seltzer C, Di Cecco G, Hurlbert AH, Guralnick R. 2023. Identifying the identifiers: How iNaturalist facilitates collaborative, research-relevant data generation and why it matters for biodiversity science. BioScience 73: 533-541.

Carman T, Hansen MBN. 2004. *The Cambridge companion to: Merleau-ponty*. Cambridge University Press.

Carney J. 2020. Thinking avant la lettre: A review of 4E cognition. *Evolutionary Studies in Imaginative Culture* 4: 77-90.

Carpenter SR, Cole JJ, Pace ML, Batt R, Brock WA, Cline T, Coloso J, Hodgson JR, Kitchell JF, Seekell DA, Smith L, Weidel B. 2011. Early warnings of regime shifts: A whole-ecosystem experiment. *Science* 332: 1079-1082.

Craver CF, Darden L. 2013. *In Search of Mechanisms: Discoveries across the Life Sciences*. University of Chicago Press; Chicago USA.

Cui H, Tejada-Lapuerta A, Brbić M, Saez-Rodriguez J, Cristea S, Goodarzi H, Lotfollahi M, Theis FJ, Wang B. 2025. Towards multimodal foundation models in molecular cell biology. Nature 640: 623-633.

Dangles O. 2023. *Climate Change on Mountains: Reviving Humboldt's Approach to Science*. Springer Nature; Cham Switzerland.

Dao A, Vu DB. 2025. AlphaMaze: Enhancing Large Language Models' Spatial Intelligence via GRPO. arXiv:2502.14669.

Daryakenari NA, De Florio M, Shukla K, Karniadakis GE. 2024. AI-Aristotle: A physics-informed framework for systems biology gray-box identification. *PLoS Computational Biology* 20: e1011916.

Dawes RM, Faust D, Meehl PE. 1989. Clinical versus actuarial judgment. *Science* 243: 1668-1674.

Devictor V, Bensaude-Vincent B. 2016. From ecological records to big data: the invention of global biodiversity. *History and Philosophy of the Life Sciences* 38: 13.

Eskuche-Keith P, Hill SL, López-López L, Rosenbaum B, Saunders RA, Tarling GA, O'Gorman EJ. 2024. Temperature alters the predator-prey size relationships and size-selectivity of Southern Ocean fish. *Nature Communications* 15: 3979.

Faure L, Mollet B, Liebermeister W, Faulon JL. 2023. A neural-mechanistic hybrid approach improving the predictive power of genome-scale metabolic models. *Nature Communications* 14: 4669.

Fogleman JC, Danielson PB. 2001. Chemical Interactions in the Cactus-Microorganism-Drosophila Model System of the Sonoran Desert. *American Zoologist* 41: 877-889.

Galiana N, Arnoldi JF, Barbier M, Acloque A, de Mazancourt C, Loreau M. 2021. Can biomass distribution across trophic levels predict trophic cascades? *Ecology Letters* 24: 464-476.

Gao S, Fang A, Huang Y, Giunchiglia V, Noori A, Schwarz JR, Ektefaie Y, Kondic J, Zitnik M. 2024. Empowering biomedical discovery with AI agents. *Cell* 187: 6125-6151.

Gao Y, Dong K, Shan C, Li D, Liu Q. 2025. Causal disentanglement for single-cell representations and controllable counterfactual generation. *Nature Communications* 16: 6775.

Ghafarollahi A, Buehler MJ. 2025. SciAgents: Automating Scientific Discovery Through Bioinspired Multi-Agent Intelligent Graph Reasoning. *Advanced Materials* 37: 2413523.

Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. 2000. Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment* 12: 19-30.

Henrys PA, Mondain-Monval TO, Jarvis SG. 2024. Adaptive sampling in ecology: Key challenges and future opportunities. *Methods in Ecology and Evolution* 15: 1483-1496.

Herzog CJ, Kays R, Ray J, Gompper M, Zielinski W, Higgins R, Tymeson M. 2007. Using Patterns in Track-Plate Footprints to Identify Individual Fishers. *The Journal of Wildlife Management* 71: 955-963.

Huebner SE, Palmer MS, Packer C. 2024. Human Supervision is Key to Achieving Accurate AI-assisted Wildlife Identifications in Camera Trap Images. *Citizen Science: Theory and Practice* 9: 10.5334/cstp.752.

Jeffery CJ. 2018. Protein moonlighting: What is it, and why is it important? *Philosophical Transactions of the Royal Society B: Biological Sciences* 373: 20160523.

Johnson WC, Alivisatos A, Smith TC, Van N, Soni V, Wallach JB, Clark NA, Fitzgerald TA, Whiteley JJ, Tan S, Sokolov A, Ando DM, Schnappinger D, Rhee KY, Aldridge BB. 2025. Integration of multi-modal measurements identifies critical mechanisms of tuberculosis drug action. *Cell Systems* 16: 101348.

Kant I. 1787. *Critique of Pure Reason (Kritik der reinen Vernunft)*. Königl. Akademie der Wissenschaften.

Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. 2021. Physics-informed machine learning. *Nature Reviews Physics* 3: 422-440.

Karpatne A, Deshwal A, Jia X, Ding W, Steinbach M, Zhang A, Kumar V. 2025. AI-enabled scientific revolution in the age of generative AI: second NSF workshop report. *npj Artificial Intelligence* 1: 18.

Kass LR. 1997. The wisdom of repugnance: why we should ban the cloning of humans. *New Republic* 216: 679-706.

Kitzes J, Beckman P, Mabee P. 2025a. Workshop on AI, Sensors, and Ecological Infrastructure (June 16-18, 2025). 2025, Pittsburgh PA, USA.

Kitzes J, Chronister L, Czarnecki C, Fiss C, Freeland-Haynes L, Goodman BD, Lapp S, Lyon RP, Nossan H, Rhinehart TA, Ruiz Guzman S, Syunkova A, Viotti L. 2025b. Integrating AI models into ecological research workflows: The case of terrestrial bioacoustics. *Methods in Ecology and Evolution* 17: 257-271.

Kleiber M. 1932. Body size and metabolism. *Hilgardia* 6: 315-353.

Leonelli S. 2016. *Data-Centric Biology*. University of Chicago Press; Chicago USA.

Leonelli S, Ankeny RA. 2015. Repertoires: How to Transform a Project into a Research Community. *BioScience* 65: 701-708.

Machamer P, Darden L, Craver CF. 2000. Thinking about mechanisms. *Philosophy of Science* 67: 1-25.

Magana Gomez PG, Kovalevskiy O. 2024. *What AlphaFold 3 struggles with*. AlphaFold: A practical guide.

Mason BM, Mesaglio T, Barratt Heitmann J, Chandler M, Chowdhury S, Gorta SBZ, Grattarola F, Groom Q, Hitchcock C, Hoskins L, Lowe SK, Marquis M, Pernat N, Shirey V, Baasanmunkh S, Callaghan CT. 2025. iNaturalist accelerates biodiversity research. *BioScience* 75: 953-965.

Merleau-Ponty M. 1969. *The Visible and the Invisible* (tr. Alphonso Lingis). Studies in Phenomenology and Existential Philosophy. Northwestern University Press.

Merleau-Ponty M. 1992. *Sense and Non-Sense* (tr. H.L. Dreyfus and P.A. Dreyfus). Studies in Phenomenology and Existential Philosophy. Northwestern University Press.

Mesaglio T, Shepherd KA, Wege JA, Barrett RL, Sauquet H, Cornwell WK. 2025. Expert identification blitz: A rapid high value approach for assessing and improving iNaturalist identification accuracy and data precision and confidence. *Plants People Planet* 7: 1469-1484.

Newen A, De Bruin L, Gallagher S. 2018. *The Oxford Handbook of 4E Cognition*. Oxford University Press; New York USA.

Nussbaum MC. 2004. *Hiding from humanity: Disgust, shame, and the law*. Princeton University Press; Princeton USA.

Paine RT. 1980. Food Webs: Linkage, Interaction Strength and Community Infrastructure. *The Journal of Animal Ecology* 49: 667-685.

Palmer MS, Huebner SE, Willi M, Fortson L, Packer C. 2021. Citizen science, computing, and conservation: How can "Crowd AI" change the way we tackle large-scale ecological challenges? *Human Computation* 8: 54-75.

Pardo LE, Bombaci S, Huebner SE, Somers MJ, Fritz H, Downs C, Guthmann A, Hetem RS, Keith M, le Roux A, Mgqatsa N, Packer C, Palmer MS, Parker DM, Peel M, Slotow R, Maartin Strauss W, Swanepoel L, Tambling C, Tsie N, Vermeulen M, Willi M, Jachowski DS, Venter JA. 2021. Snapshot Safari: A large-scale collaborative to monitor Africa's remarkable biodiversity. *South African Journal of Science* 117: 1-4.

Ripple WJ, Beschta RL. 2012. Trophic cascades in Yellowstone: The first 15years after wolf reintroduction. *Biological Conservation* 145: 205-213.

Ripple WJ, Beschta RL, Wolf C, Painter LE, Wirsing AJ. 2025. The strength of the Yellowstone trophic cascade after wolf reintroduction. *Global Ecology and Conservation* 58: e03428.

Ripple WJ, Estes JA, Schmitz OJ, Constant V, Kaylor MJ, Lenz A, Motley JL, Self KE, Taylor DS, Wolf C. 2016. What is a Trophic Cascade? *Trends in Ecology and Evolution* 31: 842-849.

Schramski JR, Dell AI, Grady JM, Sibly RM, Brown JH. 2015. Metabolic theory predicts whole-ecosystem properties. *Proceedings of the National Academy of Sciences of the United States of America* 112: 2617-2622.

Seglard D. 2003. *Nature: Course Notes from the Collège de France* (by M. Merleau-Ponty). Northwestern University Press.

Sellars W. 1956. Empiricism and the Philosophy of Mind. *Minnesota studies in the philosophy of science* 1: 253–329.

Shi Y, Xu W, Hu P. 2025. Out of distribution learning in bioinformatics: Advancements and challenges. *Briefings in Bioinformatics* 26: bbaf294.

Singh N, Bhalla N. 2020. Moonlighting Proteins. *Annual Review of Genetics* 54: 265-285.

Song P, Han P, Goodman N. 2026. Large Language Model Reasoning Failures. *Transactions on Machine Learning Research*: 10.48550/arXiv.2602.06176.

Thel L, Chamaillé-Jammes S, Keurinck L, Catala M, Packer C, Huebner SE, Bonenfant C. 2021. Can citizen science analysis of camera trap data be used to study reproduction? Lessons from Snapshot Serengeti program. *Wildlife Biology* 2021: 1-9.

Theodoris C V., Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, Hill MC, Mantineo H, Brydon EM, Zeng Z, Liu XS, Ellinor PT. 2023. Transfer learning enables predictions in network biology. *Nature* 618: 616-624.

Triplett T. 2023. Hicks on Sellars, Price, and the Myth of the Given. *Journal of the History of Analytical Philosophy* 11: 10.15173/jhap.v8i7.4270.

Van Dyck H. 2012. Changing organisms in rapidly changing anthropogenic landscapes: the significance of the 'Umwelt'-concept and functional habitat for animal conservation. *Evolutionary Applications*, 5(2), 144-153.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Zídek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S. 2022. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* 50: 439-444.

Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, Ahern W, Borst AJ, Ragotte RJ, Milles LF, Wicky BIM, Hanikel N, Pellock SJ, Courbet A, Sheffler W, Wang J,

Venkatesh P, Sappington I, Torres SV, Lauko A, De Bortoli V, Mathieu E, Ovchinnikov S, Barzilay R, Jaakkola TS, DiMaio F, Baek M, Baker D. 2023. De novo design of protein structure and function with RFdiffusion. *Nature* 620: 1089-1100.

West GB, Brown JH, Enquist BJ. 1997. A general model for the origin of allometric scaling laws in biology. *Science* 276: 122-126.

White CR, Alton LA, Bywater CL, Lombardi EJ, Marshall DJ. 2022. Metabolic scaling is the product of life-history optimization. *Science* 377: 834-839.

Wilson M. 2002. Six views of embodied cognition. *Psychonomic Bulletin & Review* 9(4): 625-636.

Wolpert L. 1994. *The Unnatural Nature of Science*. Harvard University Press; Cambridge USA.

Wu W, Mao S, Zhang Y, Xia Y, Dong L, Cui L, Wei F. 2024. Mind's eye of LLMs: visualization-of-thought elicits spatial reasoning in large language models. *Advances in Neural Information Processing Systems* 37: 90277-90317.

Xu C, Silliman BR, Chen J, Li X, Thomsen MS, Zhang Q, Lee J, Lefcheck JS, Daleo P, Hughes BB, Jones HP, Wang R, Wang S, Smith CS, Xi X, Altieri AH, de Koppel J van, Palmer TM, Liu L, Wu J, Li B, He Q. 2023. Herbivory limits success of vegetation restoration globally. *Science* 382: 589-594.

Xu H, Usuyama N, Bagga J, Zhang S, Rao R, Naumann T, Wong C, Gero Z, González J, Gu Y, Xu Y, Wei M, Wang W, Ma S, Wei F, Yang J, Li C, Gao J, Rosemon J, Bower T, Lee S, Weerasinghe R, Wright BJ, Robicsek A, Piening B, Bifulco C, Wang S, Poon H. 2024. A whole-slide foundation model for digital pathology from real-world data. *Nature* 630: 181-188.