

AI Biology: a new discipline for discovering principles of life and disease emergence

Chao Yan^{1#}, Zhen Zhou^{1,2,3#}, Botao Li^{1,2,3}, Zefeng Wang⁴, Dangsheng Li⁵, Chen-Yu Zhang^{*1,2,3}

<https://doi.org/10.15302/vita.2026.09.0075>

Over the past two decades, life sciences have been transformed by genome editing, single-cell multi-omics, cryo-EM, and high-resolution imaging¹. Yet a persistent gap remains between discovery and translation. This gap reflects several enduring bottlenecks: difficulty in integrating heterogeneous high-dimensional data, the limits of reductionism in explaining non-linear multiscale phenomena, fragmentation across biological scales, and the slow prediction-validation loop between theory and experiment. These issues arise not merely from insufficient data or tools, but from the absence of a discipline-level framework that can integrate multimodal information, capture emergence, model multiscale dynamics, and continuously close the loop between computation and perturbation.

In parallel, artificial intelligence (AI) has achieved breakthroughs in representation learning and generative modeling, with transformer architectures enabling powerful sequence-based learning². In biology, AI has already demonstrated its ability to capture latent regularities, most prominently in protein structure prediction³. These advances suggest that AI can be more than an auxiliary analytical tool: it can become a cognitive framework and theory-construction engine for biology. This motivates the establishment of AI Biology (AIBIO) as a new discipline.

CORE CONCEPT AND THEORETICAL FOUNDATION OF AI BIOLOGY

Definition of AI Biology: life emergence and biological wisdom

AIBIO is a new interdisciplinary discipline that studies the laws of life emergence and biological wisdom. AIBIO is oriented toward the central “Why” questions in life sciences: why life can form, why life activities are organized and stabilized in certain ways, and why universal motifs recur across scales.

Here, life emergence is not a generic synonym for complexity, nor is it confined to organ-level phenomena. It refers to a pan-scale process spanning molecules, cells, tissues, organisms, populations, and ecosystems, in which system-level structures, dynamical regimes, and functions arise from local interactions, feedback, and spatiotemporal coupling. Emergent properties are not present in the same form at lower levels and cannot be fully explained by linear superposition or single-component mechanisms, echoing the general principle

that “more is different”⁴. Crucially, life emergence is shaped by cross-scale coupling: bottom-up interactions generate new collective behaviors, while top-down constraints (e.g., tissue architecture, physiology, environment) select and stabilize particular regimes.

In AIBIO, biological wisdom is used in a strictly non-anthropomorphic sense. It denotes principle-level regularities: compact, transferable and falsifiable mathematical constraints that explain why living systems repeatedly achieve robustness, homeostasis, adaptation, evolvability and recurrent coding architectures under physical, chemical and evolutionary constraints⁵. A candidate “wisdom-level” statement must (i) generalize beyond a single dataset, model, or biological context; (ii) specify its domain of validity; and (iii) predict the conditions under which it would fail. In this sense, biological wisdom is the principle-level answer to “Why” — beyond, but inseparable from, mechanism-level “How”.

Emergence, life emergence, and disease emergence: what exactly is studied

To sharpen AIBIO’s theoretical target, it is useful to distinguish three related concepts.

Emergence refers to macroscopic properties that arise from non-linear interactions and are absent at the component level⁴. AIBIO treats emergence not as a descriptive label but as a scientific object: identifying minimal organizing variables, coupling rules, and stability conditions that make emergent properties predictable and falsifiable.

Life emergence is a special class of emergence in which hierarchical organization yields novel structures, dynamical regimes, and functions across biological scales and stabilizes them through cross-scale constraints. Accordingly, AIBIO studies (i) emergent structures (modules, networks, spatial patterns); (ii) emergent dynamics (state transitions, attractors, oscillations, tipping points); (iii) emergent functions (robustness, homeostasis, adaptive control); and (iv) cross-scale universals that can be expressed mathematically and tested experimentally.

Disease emergence follows the same logic. AIBIO views disease as an emergent, multi-level phenotype rather than a single-layer defect. A practical principle-level description is: disease reflects a living system entering a maladaptive regime under specific spatiotemporal contexts, shaped by interac-

1. Nanjing Drum Tower Hospital Center of Molecular Diagnostic and Therapy, State Key Laboratory of Pharmaceutical Biotechnology, Jiangsu Engineering Research Center for MicroRNA Biology and Biotechnology, NJU Advanced Institute of Life Sciences (NAILS), School of Life Sciences, Nanjing University, Nanjing, Jiangsu, China. 2. Research Unit of Extracellular RNA, Chinese Academy of Medical Sciences, Nanjing, Jiangsu, China. 3. Tianfu Jincheng Laboratory, Chengdu, Sichuan, China. 4. Department of Systems Biology, School of Life Sciences, Guangming Advanced Research Institute, Southern University of Science and Technology, Shenzhen, Guangdong, China. 5. Shanghai Academy of Natural Sciences, Shanghai, China. [#]These authors contributed equally. *Correspondence: Chen-Yu Zhang (cyzhang@nju.edu.cn)

Received: April 1, 2026; Accepted: September 26, 2026; Published: October 22, 2026

tions among molecular programs, cell states, tissue microenvironments, organismal physiology, and external ecology. This aligns with disease-as-network and disease-as-state-transition perspectives: diseases often arise from network perturbations⁶ and can be framed as transitions between dynamical “attractor” states in complex regulatory systems⁷, including critical transitions where small perturbations trigger large shifts. Under this view, intervention is not only “fixing a molecule” but steering a multiscale system back from a maladaptive regime toward adaptive stability. For clinical translation, this view points to concrete actionable outputs: (i) early-warning indicators of regime transitions; (ii) regime-defining biomarker panels rather than isolated single markers; (iii) combination perturbation strategies that shift system state rather than merely suppress one molecule; and (iv) patient-specific digital twins for predicting therapeutic response. AIBIO thus extends, rather than replaces, target-centric thinking by embedding molecular targets within a multiscale state-space and intervention design.

DIKW as the operational logic: from data to unified laws (“Why”)

AIBIO can be operationally organized by the DIKW hierarchy — Data → Information → Knowledge → Wisdom — which is especially useful for separating prediction from principle.

Data are multimodal, multiscale measurements (sequence, omics, spatial profiles, imaging, physiology, behaviors, environment) plus perturbations. A central AI task is representation learning: converting heterogeneous signals into computational embeddings.

Information emerges when AI extracts structure and patterns (modules, networks, trajectories, spatial organization). Network-focused machine learning provides key methodological routes from data to structured biological relations.

Knowledge forms when information is organized into predictive and actionable models, ideally with causal content. This includes models that propose perturbations and predict their effects, enabling experimentally falsifiable hypotheses.

Wisdom is the discipline-defining goal of AIBIO: formalizable, generalizable, and refutable unifying principles — mathematical laws — that could answer “Why” across systems, including both life and disease emergence. By institutionalizing this DIKW ascent, AIBIO aims to move biology from a predominantly mechanism-centered science (“How”) toward a principle-centered science (“Why”), while keeping experimental falsification as the ultimate standard (Fig. 1).

To avoid conceptual inflation, an AIBIO Wisdom-level claim must be operationally falsifiable rather than rhetorically appealing. A candidate principle should specify the order parameters or collective variables that are measured or inferred, together with an explicit domain of validity — the organisms, cell types, diseases, timescales or environments to which it applies. It should be expressed in a definite mathematical form, such as an equality, inequality, scaling relation, conservation-like constraint, trade-off, attractor structure or transition rule, and should be evaluated against well-defined null models obtained from randomization, shuffled interactions or matched marginals. Critically, it must yield concrete perturbation predictions — what should happen when a component, context, or constraint is altered — and corre-

sponding failure criteria that specify which empirical result would refute or restrict the principle. Under this standard, a statement such as “disease is an attractor” is not yet a principle; it becomes one only when its state variables, regime boundaries, perturbation responses and refutation conditions are explicitly stated and experimentally vulnerable.

WHY ESTABLISH AI BIOLOGY AS A DISCIPLINE: NECESSITY AND IMPORTANCE

Modern biology increasingly faces problems where local mechanisms alone do not yield general explanations. High-dimensional multimodal datasets capture biological reality more comprehensively than ever, yet they also increase the risk of accumulating fragmented mechanistic stories without converging to unifying principles. Meanwhile, translation requires predictive, robust models that generalize across individuals, contexts, and scales — precisely where emergence and cross-scale coupling dominate.

AIBIO is necessary because it provides a discipline-level framework to (i) integrate multimodal/multiscale evidence; (ii) discover organizing variables and governing rules from complex observations; (iii) formulate falsifiable principles; and (iv) close the loop through perturbation and validation. This ambition echoes early work on inferring governing rules from observations in non-linear dynamical systems⁸, but now becomes feasible at biological scale due to modern AI and measurement technologies.

Three concrete questions illustrate directions that AIBIO seeks to address beyond existing frameworks. First, the architecture of the genetic code: why does canonical translation use a 20-amino-acid core and triplet codons? The classical “frozen accident” view emphasizes historical contingency⁹, but quantitative analyses suggest the standard code occupies a privileged region of an error-minimizing landscape¹⁰. AIBIO may help integrate chemical constraints, codon-mutation neighborhoods, translational errors, and synthetic-code design to test whether the standard code reflects a structured, navigable coding landscape. Second, the geometry of evolution: is adaptation only random variation followed by selection, or do adaptive trajectories repeatedly flow through constrained channels, like water through pre-carved riverbeds? Canalization and fitness-landscape ideas already suggest that variation operates within structured spaces¹¹. AIBIO may help integrate genotype-phenotype maps, ecological context, and evolutionary trajectories to test whether adaptive flows obey identifiable statistical, geometric or dynamical regularities. Third, the persistence of disease regimes: why are cancers, metabolic disorders, and neurodegeneration so often stable and difficult to reverse? Dynamical network biomarker theory shows disease progression involves detectable pre-transition states¹²; AIBIO may extend this by coupling representation learning with perturbation design and therapeutic state-steering.

Its importance is both scientific and practical. Scientifically, AIBIO creates a home for principle-oriented theory building about life and disease emergence. Practically, it enables a new research engine: models continuously updated by data and challenged by experiments, accelerating discovery-to-translation cycles in medicine, synthetic biology, and environmental systems.

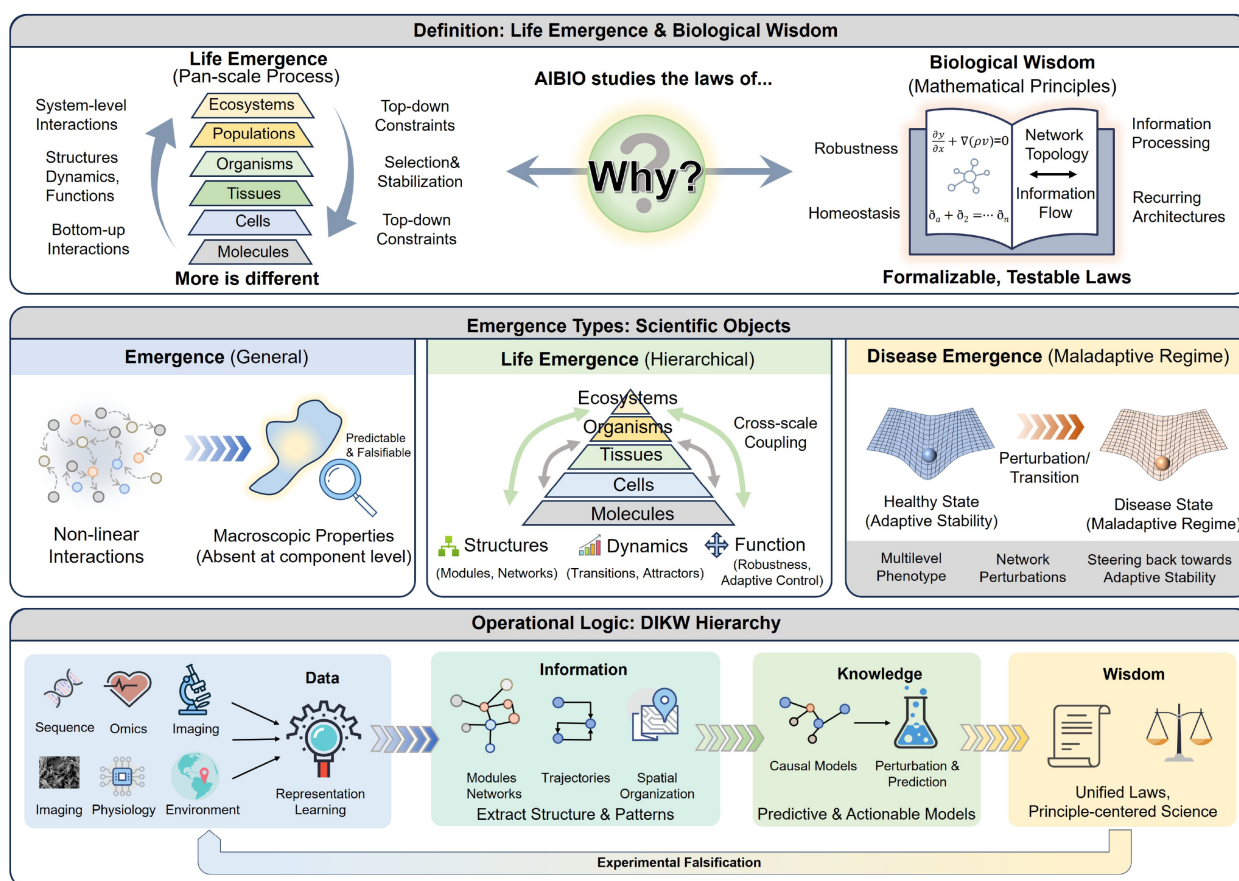


Fig. 1 Core concept and theoretical foundation of AIBIO. AIBIO drives a paradigm shift in the life sciences from mechanism-centered observation (“How”) to principle-centered theory (“Why”). The discipline is structurally organized by the Data-Information-Knowledge-Wisdom (DIKW) hierarchy. Heterogeneous, pan-scale Data are processed via AI representation learning to extract structured Information, such as network motifs and spatial patterns. This information is subsequently transformed into predictive and causal Knowledge through computational modeling and targeted perturbations. Guided by a continuous closed-loop cycle of computational prediction and experimental falsification, this progressive ascent ultimately yields Wisdom: formalizable, mathematical laws that govern the multiscale emergence of life and disease. At the Wisdom layer, symbolic expressions denote candidate cross-scale invariants and constraints — for example, order parameters Φ that summarize collective state, transition rules of the form $\frac{d\Phi}{dt} = F(\Phi; \Theta)$ describing regime dynamics, and conservation- or trade-off-like relations among coupled variables — each of which must be associated with measurable quantities, a stated domain of validity and falsifiable perturbation predictions.

AIBIO vs AI FOR BIOLOGY (AND BOUNDARIES WITH ADJACENT DISCIPLINES)

AI for Biology applies AI as a tool to improve established biological workflows — e.g., image segmentation, phenotype recognition, sequence analysis, or task-specific prediction. These advances are valuable, but they typically leave the conceptual framework of biology unchanged: AI enhances efficiency, while biology remains centered on local mechanism accumulation.

AIBIO differs in its discipline-level outputs and epistemic targets. It uses AI as a cognitive and modeling framework to reconstruct biology toward “Why”, and uses biological perspectives to interrogate what robust learning and adaptation should mean in complex systems. The expected outputs are not only tools, but also new concepts, paradigms, and unifying laws that are cross-scale and experimentally falsifiable.

This clarifies boundaries with adjacent fields. Systems biology emphasizes integrative mechanistic thinking and cross-level modeling, often with strong prior assumptions and explicit equations. Bioinformatics emphasizes data manage-

ment, annotation, and statistical analysis infrastructure¹³. AIBIO incorporates both but is defined by DIKW-driven ascent to principles, emergence as a central target (including disease emergence), and closed-loop validation as a disciplinary norm. A specific operational distinction is that systems biology typically begins with known variables, defined pathways and explicitly specified equations, whereas AIBIO is required precisely when the relevant variables, scales and coupling rules are unknown *a priori* and must be induced from multimodal perturbation data. AIBIO therefore does not separate cleanly from systems biology; it is a higher-level synthesis that adds three discipline-defining requirements: principle-oriented outputs, cross-scale emergence as a central object, and closed-loop experimental falsification as a norm (Fig. 2).

RESEARCH SCENARIOS OF AIBIO AND THE TRANSFORMATIONS THEY ENABLE

Multimodal integration: toward principle-bearing biological atlases

AIBIO requires unified representations that connect

sequences, structures, single-cell states, spatial context, physiology, and environment. Integrative single-cell analysis exemplifies the alignment problem across datasets and modalities, while deep learning has expanded measurable phenotype spaces in cellular imaging¹⁴. Multimodal integration becomes transformative when atlases are no longer descriptive catalogs but predictive objects that expose cross-scale constraints and candidate organizing variables¹⁵.

Explainable and causal modeling: from correlations to falsifiable “Why” hypotheses

Principle-seeking biology requires models whose explanations survive perturbation. Feature attribution methods (e.g., SHAP and integrated gradients) provide useful starting points but do not by themselves establish causality. Perturbation-rich designs — including pooled CRISPR screens and single-cell perturbation profiling such as Perturb-seq¹⁶ and multiplexed CRISPR readouts¹⁷ — provide a more direct route from prediction toward causal knowledge, and AIBIO ultimately demands such perturbation-consistent causal accounts. For example, inferred cell-cell communication programs from single-cell data can suggest experimentally testable intervention points in tumors¹⁸. In the AIBIO framing, the goal is not only to predict outcomes but to identify the minimal causal structure that explains why a system stabilizes in a particular regime (healthy or diseased).

Digital twins and closed-loop science: steering systems across health-disease regimes

Digital twins — virtual counterparts continuously updated by real data — are becoming a general paradigm in complex systems¹⁹. In biology, whole-cell computational modeling points toward integrated genotype-to-phenotype engines²⁰. AIBIO extends this direction by coupling digital twins with algorithmic planning and learning. Reinforcement learning and model-based planning demonstrate how systems can improve decisions through iterative simulation and feedback²¹, providing a template for AI-guided experiment selection. A concrete AIBIO closed-loop cycle contains five components: (i) a model predicts system state together with calibrated uncertainty; (ii) an active-learning or Bayesian-optimization algorithm selects the maximally informative perturbation; (iii) the perturbation is implemented experimentally — for example by CRISPR knockout, RNAi knockdown, drug treatment or environmental change; (iv) multimodal readouts (single-cell, spatial, imaging, functional assays) update the model and its uncertainty estimates; and (v) candidate principles are retained, revised, or rejected against pre-specified failure criteria. The loop is therefore not just “predict-test-update” but a structured pipeline in which the model is evaluated by its ability to design perturbations that change biological regimes in predicted ways. In medicine, this offers a route to interpret disease as a multiscale maladaptive regime and to rationally design interventions that steer systems back toward adaptive stability.

Generative biology: design as a stringent test of understanding

AIBIO treats generation not only as engineering but as epistemic validation: if a model truly captures biological

constraints, it may in principle be able to generate viable candidates consistent with those constraints. Foundational generative modeling ideas have evolved into powerful tools for molecular docking and design, including diffusion-based de novo protein design (RDiffusion)²² and learned-representation sequence design (ProteinMPNN)²³. In AIBIO, such generative capacity is interpreted as a stress test for whether the inferred “rules” reflect real biological constraints: successful and experimentally functional generation is positive evidence that the model has captured part of the rule space governing viability or specificity, while systematic generative failure flags where the inferred principles are incomplete and indicates which constraints must be added next.

An RNA-centered illustration: RNA programs as cross-scale control variables for emergence and disease

RNA biology is a particularly strong testbed for AIBIO because it links sequence, structure, dynamics (processing, decay, localization), and regulation, while offering abundant perturbation options. In a conventional “How”-centered framing, RNA studies often accumulate mechanisms (factors, motifs, modifications). In the AIBIO “Why” framing, RNA becomes a window into deeper questions: why certain regulatory architectures recur, why specific dynamical regimes (buffering, switching, hysteresis) are stabilized, and how noisy molecular components yield reliable cell-state control.

Along the DIKW path, transcriptomic and spatial measurements provide high-dimensional Data; AI extracts information as latent RNA programs, trajectories, and regulatory modules; combined with perturbations, models can build Knowledge that predicts and controls state transitions²⁴. The AIBIO target is Wisdom: falsifiable principles describing why RNA-based regulation stabilizes particular emergent regimes, both adaptive and maladaptive. While no full AIBIO closed loop has yet been experimentally demonstrated for an RNA system, the following scenario illustrates what such a cycle could entail. As a concrete instantiation, an AIBIO model could propose a panel of RNAi or antisense perturbations predicted to displace a diseased single-cell state distribution toward a healthier attractor; single-cell and functional readouts would then test whether the predicted state transition occurs, and discrepancies would feed back into model revision. This illustrates how RNA expertise could contribute to principle discovery: RNA is both information-rich and perturbation-accessible and intrinsically multiscale, enabling the closed-loop falsification that AIBIO requires²⁵.

SUMMARY AND OUTLOOK

AIBIO is being established to answer life science “Why” questions through a discipline-level framework: studying life emergence and biological wisdom as formalizable, testable mathematical principles. AIBIO emphasizes: (i) pan-scale life emergence from molecules to ecosystems; (ii) disease as an emergent multi-level maladaptive regime under spatiotemporal context; and (iii) DIKW ascent from multimodal data to unifying laws. Methodologically, it relies on multimodal integration, explainable and causal modeling, digital twins with closed-loop validation, and generative testing/design.

Looking ahead, AIBIO must confront three core scientific challenges. First, mathematically measuring emergence: the

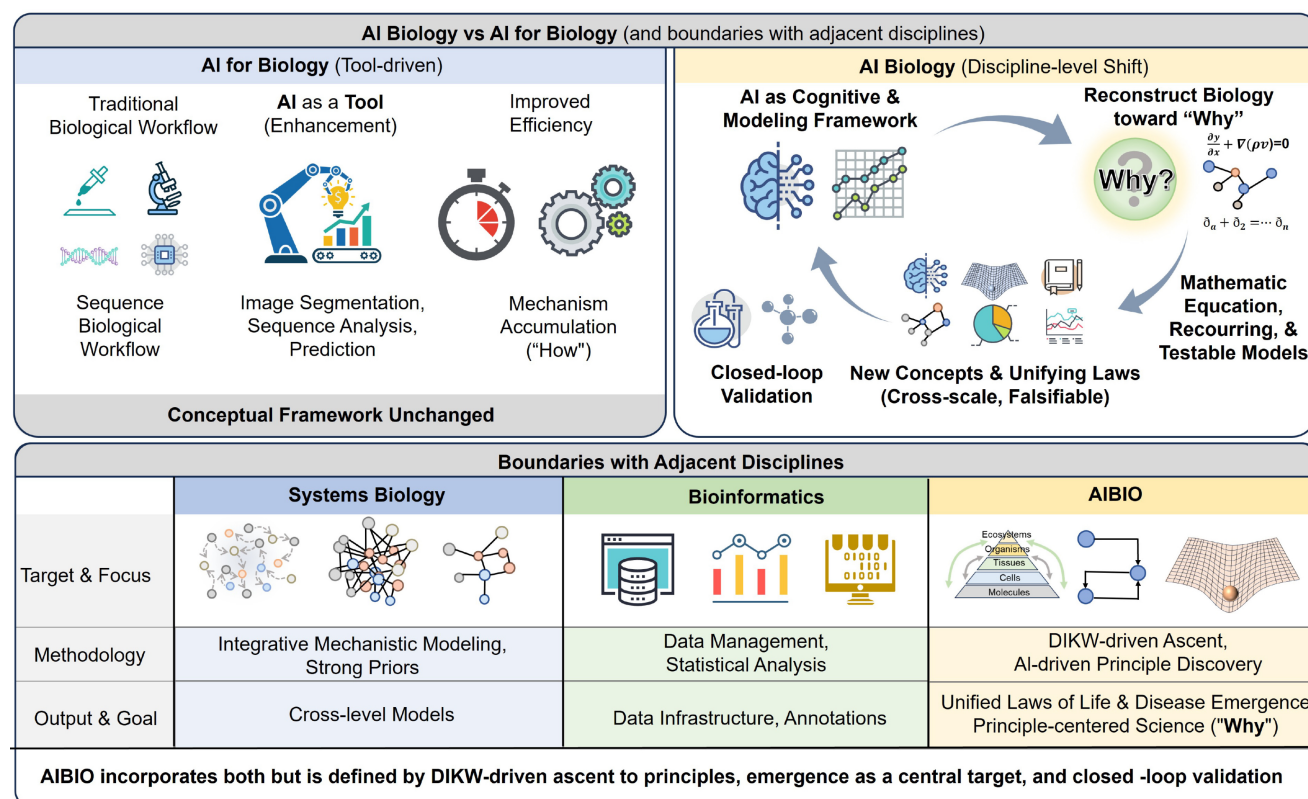


Fig. 2 AIBIO vs AI for biology and adjacent disciplines. A comparison of AIBIO with established fields highlights its distinct role as a cognitive framework and theory-construction engine. While bioinformatics provides essential data management infrastructure and systems biology emphasizes mechanistic cross-level modeling, AI for biology primarily applies AI as an auxiliary tool to optimize traditional, mechanism-accumulating workflows. AIBIO is shown as building upon, rather than replacing, these adjacent fields: it inherits data infrastructure from bioinformatics, quantitative modeling from computational biology, integrative mechanistic frameworks from systems biology, and task-level prediction and design capabilities from AI for biology. In contrast, AIBIO represents a distinct, discipline-level paradigm. By integrating the computational and integrative strengths of adjacent fields, AIBIO is uniquely defined by its DIKW-driven ascent and closed-loop experimental validation, aiming to discover the unifying, cross-scale governing laws ("Why") underlying biological emergence.

field needs quantitative order parameters that distinguish genuine cross-scale emergence from high-dimensional correlation, are robust to data modality, sampling depth and model architecture, and are linked to perturbation outcomes rather than left as descriptive statistics. Second, making causal AI biologically interpretable: causal discovery in biology is hard because perturbations are incomplete, contexts vary and feedback is pervasive; AIBIO models must combine causal inference with mechanistic biological constraints, so that interpretability includes perturbation-consistent causal structure and not only feature attribution. Third, linking scales without collapsing them: molecular, cellular, tissue, organismal and ecological scales differ in variables and timescales, and AIBIO must avoid both reduction to molecular lists and vague holistic narratives — hierarchical latent models, graph neural operators, causal representation learning and multi-fidelity simulators offer practical routes. As aspirational field-level targets, two field-level milestones can anchor progress: within five years, perturbation-rich digital twins for selected systems — cell-fate transitions, organoids, immune activation or specific disease-state transitions — may become capable of predicting and experimentally steering state changes; within ten years, the field may begin to establish benchmarked cross-scale principles and order parameters

for several biological processes, enabling AI-designed interventions that shift maladaptive disease regimes and AI-guided design of synthetic biological systems. If achieved, AIBIO will not only accelerate discovery and translation but also reshape biology into a principle-seeking science of emergence in life and disease.

ACKNOWLEDGMENT

This work was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS-2021-12M-5-015), the National Natural Science Foundation of China (32341007, 91857000, 32022015), and the "Open Competition to Select the Best Candidates" Key Technology Program for Nucleic Acid Drugs of NTICB (2022HS02008).

AUTHOR CONTRIBUTIONS

C.-Y.Z. conceptualized the core idea of the article and guided the manuscript writing. C.Y. and Z.Z. wrote the manuscript and revised the manuscript. B.L. prepared the figures. Z.W. and D.L. provided valuable suggestions during the development of the manuscript and assisted in revising the manuscript. All authors reviewed the manuscript, approved the final version, and agreed to be accountable for all aspects of the work.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

- Luecken, M.D. & Theis, F.J. *Mol. Syst. Biol.* **15**, e8746 (2019).
- Vaswani, A. et al. <https://doi.org/10.48550/arXiv.1706.03762> (2023).
- Jumper, J. et al. *Nature* **596**, 583–589 (2021).
- Anderson, P.W. *Science* **177**, 393–396 (1972).
- Kitano, H. *Nat. Rev. Genet.* **5**, 826–837 (2004).
- Barabási, A.L., Gulbahce, N. & Loscalzo, J. *Nat. Rev. Genet.* **12**, 56–68 (2011).
- Huang, S. *Cancer Metastasis Rev.* **32**, 423–448 (2013).
- Brunton, S.L., Proctor, J.L. & Kutz, J.N. *Proc. Natl. Acad. Sci. USA* **113**, 3932–3937 (2016).
- Crick, F.H.C. *J. Mol. Biol.* **38**, 367–379 (1968).
- Freeland, S.J. & Hurst, L.D. *J. Mol. Evol.* **47**, 238–248 (1998).
- Waddington, C.H. <https://doi.org/10.4324/9781315765471> (1957).
- Liu, R., Wang, X.D., Aihara, K. & Chen, L.N. *Med. Res. Rev.* **34**, 455–478 (2014).
- Stephens, Z.D. et al. *PLoS Biol.* **13**, e1002195 (2015).
- Christiansen, E.M. et al. *Cell* **173**, 792–803.e19 (2018).
- Moses, L. & Pachter, L. *Nat. Methods* **19**, 534–546 (2022).
- Dixit, A. et al. *Cell* **167**, 1853–1866.e17 (2016).
- Adamson, B. et al. *Cell* **167**, 1867–1882.e21 (2016).
- Efremova, M., Vento-Tormo, M., Teichmann, S.A. & Vento-Tormo, R. *Nat. Protoc.* **15**, 1484–1506 (2020).
- Tao, F., Qi, Q.L., Wang, L.H. & Nee, A.Y.C. *Engineering* **5**, 653–661 (2019).
- Karr, J.R. et al. *Cell* **150**, 389–401 (2012).
- Jaderberg, M. et al. *Science* **364**, 859–865 (2019).
- Watson, J.L. et al. *Nature* **620**, 1089–1100 (2023).
- Dauparas, J. et al. *Science* **378**, 49–56 (2022).
- Norman, T.M. et al. *Science* **365**, 786–793 (2019).
- Cech, T.R. & Steitz, J.A. *Cell* **157**, 77–94 (2014).

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Chen-Yu Zhang.

Reprints and permission information is available at <https://www.vita-journal.com/>.

© The Author(s) 2026. Published by Higher Education Press. This is an Open Access article distributed under the terms of the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).