Preprints.org

Concept Paper

Not peer-reviewed version

World Models for Biomedicine: Prediction Is the Means, Selection Is the End

Jianghui Xiong *

Posted Date: 24 September 2026

Keywords: world model; spatial selection; biomedicine; harness; predictability; steerability; longevity medicine; life world model; Capomics; mIC vector; intervention response; DNA methylation; N-of-1; virtual patients

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 September 2026 doi:10.20944/preprints202609.2101.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Concept Paper

World Models for Biomedicine: Prediction Is the Means, Selection Is the End

Jianghui Xiong

DeepoMe Limited, Beijing, China; jianghui@deepome.com or xiongjh77@163.com

Abstract

Biomedical world models are usually built to predict the next state. For deciding interventions, however, a prediction matters only when it improves selection: which action to take, for this person, now. We therefore argue for predictability before prediction - learn where prediction works before predicting. This perspective proposes the route; two published studies anchor it, and the rest is scheduled work. Step one: build a state map. Health is represented as 332 named gene modules, each with an interpretable capability score - a compressed, readable map of where a person stands. Step two: learn the dependency structure, which modules regulate which. An open 332 x 1,916 drug-module matrix already records what each intervention is expected to move; an initial hallmark-to-organ dependency map is under construction. Step three: test the map against paired before-and-after data, the data type the Cell Perspective names as the field's binding constraint. Each response will be read through two lenses: a single-score clock and the module map. Which lens can read intervention response is exactly what the agenda will decide. The negative results will mark where prediction is not supported. Selection then has a working definition: act where the map and the data agree, and stay out where they do not. The deliverable is not a bigger predictor but a harness for state evolution. It is the biological counterpart of what turns a raw language model into a useful, bounded system; this is what the program's term steerability has always meant. In one line: prediction is the means, selection is the end. The contribution is not a validated clinical decision system. It is a proposed, falsifiable route to intervention-oriented biomedical world models before longitudinal data are abundant.

Keywords: world model; spatial selection; biomedicine; harness; predictability; steerability; longevity medicine; life world model; Capomics; mIC vector; intervention response; DNA methylation; N-of-1; virtual patients

1. Introduction: The Gap in Longevity Medicine

Longevity medicine has solved detection faster than it has solved decision-making. Epigenetic clocks [1,2], multi-omic panels, and organ-function tests now measure biological state routinely. But a clinic that measures ten clients with the same instruments will get ten different intervention plans, none traceable to the measurements. A typical client runs five to ten concurrent interventions - supplements, hormones, drugs, training. After one year, no system can say which one worked, which one was wasted, and which one did harm. The N-of-1 paradigm has been proposed as the remedy, with an ecosystem vision for single-person analytics [3]; but without a state representation that can read response, the vision remains an aspiration.

The deeper issue predates longevity clinics: clinical care is interventional by nature. Every treatment choice is a counterfactual question [4] - what happens to *this* person if we act, versus if we do not - yet most biomedical AI answers only factual ones. A predictor that cannot read its own failure conditions cannot be used in that loop, whatever its benchmark accuracy.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 September 2026 doi:10.20944/preprints202609.2101.v1

1.1. What This Perspective Does

This perspective proposes a route; it is not a literature survey and not a delivery report. Two published studies anchor the route: a steerable N-of-1 reasoning engine, and a public 332 x 1,916 drug-module benchmark (Section 4 names them). The further steps are scheduled work, to be reported as they mature. The idea comes first; the studies are how it will be tested.

The title states the inversion plainly: for health intervention, prediction is the means, selection is the end. A dominant practical response has been to seek more longitudinal data, larger models, and more accurate next-state prediction. This perspective transposes the problem instead: temporal prediction by spatial selection - learning the dependency structure that constrains which transitions are worth testing (Fig. 1). Two further commitments follow. First, the order of research: before predicting, study predictability. Whether responses can be read at all - in which representation, under which conditions - is itself the research object, and its negative results bound the usable region. Second, the nature of the deliverable: not a bigger predictor but a *harness for state evolution*. Two working pictures carry the whole idea. The harness analogy: a raw model can go anywhere, and the engineered harness makes it useful and bounded. The shepherd principle: a wise shepherd does not map the entire grassland before moving the flock. He studies where the water and grass are, where the wolves are, and plans a safe path. The search space is narrowed by principled constraint, not exhaustive mapping. The claim concerns what to build and in what order. It is not about the physical controllability of organisms, and it is not a declaration of completion (Table 2 states, for every claim, how it will be tested).

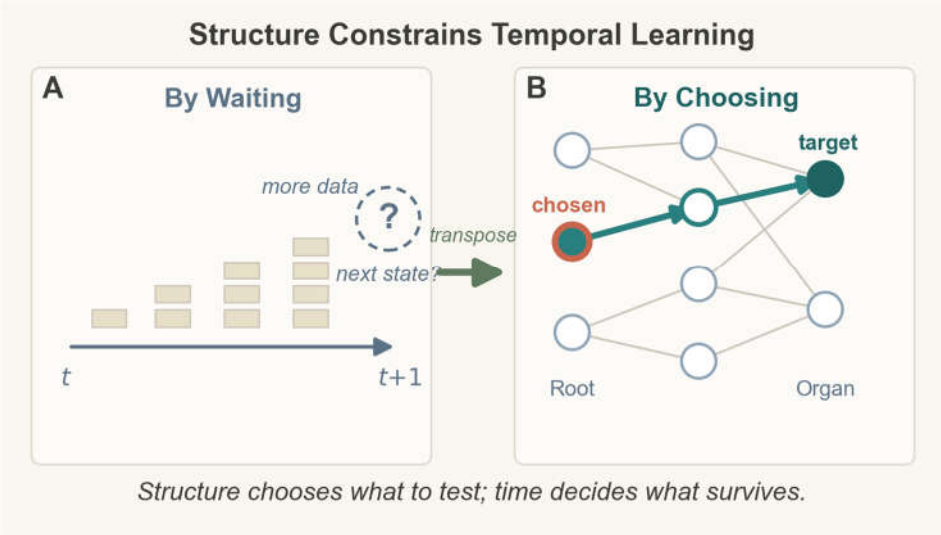


Figure 1. Structure constrains temporal learning. (A) By waiting, the default route: data accumulate along the timeline, and the next state remains a question mark - the region the field names as its binding constraint. (B) By choosing, the transposition proposed here: a candidate upstream node and its downstream target define a directional transition hypothesis, whose support is assessed with paired observations. Structure does not replace longitudinal evidence. It narrows the candidate transition hypotheses that scarce longitudinal observations are asked to test; those observations then support, revise, or reject the proposed structure.

2. The Field Has Clarified the Destination

The world-model idea entered biomedicine from AI, where model-based agents act by predicting the consequences of their own actions [5,6]. Existing biomedical world models fall into four routes: deductive constraint from first principles, generative simulators trained on data, rubric-style roadmaps for what medical world models should progressively achieve [7,8,9], and organization-mimicking agent systems that orchestrate multi-source evidence for pipeline-level decisions. None is hypothetical. Tumor evolution has been simulated under competing treatment

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 September 2026 doi:10.20944/preprints202609.2101.v1

plans [10]. VCWorld couples biological knowledge with language-model reasoning to predict drug response [11]. Evaluation frameworks now judge virtual cells by whether predictions survive intervention and counterfactual tests [12]. A major pharma-AI partnership has moved virtual-cell models into oncology drug development [13]. And a virtual biotech of tens of thousands of agents now prioritizes targets and dissects failed trials under a chief-science-officer orchestration [14]. The newest of these selects at pipeline scale - which target, which modality; the route argued here selects at person scale - which intervention, for this state. Both are selection problems; the difference is the state representation the selection runs on.

The definitional frontier has consolidated at journal level. A Perspective in *Cell* defines a biomedical world model functionally: a next-state predictor, given a current state, a specified action, and a time interval [15]. The same document lays out data, modeling, and validation routes across cells, embryos, and patients. Two features of that document matter here. It is a framework and roadmap, not a delivered model. Its authors state that no comprehensively validated general biomedical world model yet exists, and they nominate the entry tests any candidate must pass. Section 5 reads the route against them, one by one. And of its three showcase scales, it singles out the patient scale as the hardest: treatment is never randomly assigned in records, and longitudinal, intervention-linked data are the binding constraint.

The program reviewed here enters exactly at that constraint, not with a rival definition. Its empirical core is a growing collection of paired before-and-after observations at the individual human scale, and its layers can be read against the Perspective's own tests. What the field's documents leave open is the practical implementation: how any one program can work toward that destination with the data it already has. That unresolved question - not a rival definition - is the space this perspective occupies.

3. The Framework: A Harness for State Evolution

3.1. Predictability and Steerability, Defined Apart

Because this perspective orders predictability before prediction, both terms need operational definitions. *Predictability is treated as a research object*: whether, in a given state representation and under given intervention and data conditions, responses can be reliably read. It is representation-dependent and conditional, and its negative results are themselves findings. *Steerability is the application built on predictability knowledge*: intervening with a declared expected response, reading the deviation, attributing it to a named component, and revising within an evaluated cost. The two are distinct: a self-diagnosing model is not thereby steerable, and a prediction engine without boundary knowledge is neither.

The framework moves the field-level gap to the level of the model's parts. Five constraint checkpoints form a closed loop [7]. **CP1**, state representation: module-level intrinsic capabilities - the mIC vector, one interpretable score per named module, not a single number. **CP2**, quantification: molecular measurements become a capability index per module. **CP3**, intervention-response semantics: all inputs share one encoding space. **CP4**, state-transition hypothesis: state at one time projects into the next under a declared action representation, time interval, and explicit assumptions. **CP5**, five-gate inspection (State, Input, Response, change in mIC, Phenotype): a failed prediction becomes a diagnosable signal rather than an opaque negative result. The resulting shift is from a "what-if" simulator toward a quality-controlled "why-not" steering system.

3.2. The Harness, Stated Plainly

Read together, the checkpoints are a *harness for state evolution* (Fig. 2) - constrained, inspectable health-state transitions driven through harness engineering, the biological counterpart of what turns a raw large language model into a useful, bounded system [16]. The harness adds four components on a suitable state representation. **Constraints**: the declared causal scaffold. Two working postulates carry it: life as an ensemble of adaptive capacities, and module response patterns as the common

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 September 2026 doi:10.20944/preprints202609.2101.v1

encoding space. They yield a three-layer scaffold, declared before modeling, so that data calibrate it and can falsify it. **Intervention semantics**: which levers exist and what they are expected to move. **Objectives**: which capabilities to maintain or restore. **Boundaries**: where prediction is not supported, from predictability research. One disanalogy should be stated plainly: the harness constrains the model's inferences, not the organism. It is an engineering discipline on the predictor, not a mechanism claimed to steer biology itself. *Steerable*, in this program's vocabulary, means exactly this: a model equipped with a harness. Readers arriving from AI engineering will recognize the concept under either name.

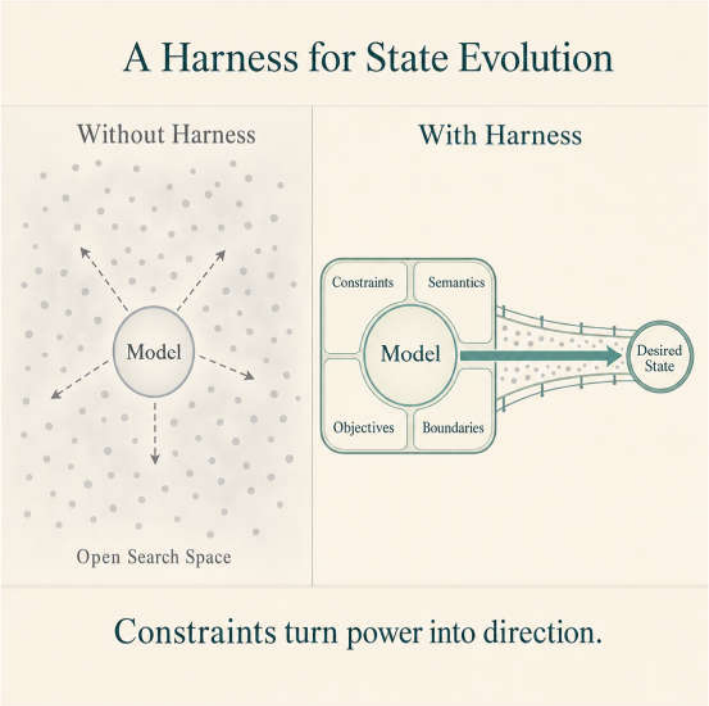


Figure 2. Two working pictures of one principle. Left panel, without a harness: a raw next-state predictor, like a raw large language model, can go anywhere - a liability in medicine. Right panel, with a harness: the engineered harness wraps the model with constraints, intervention semantics, objectives, and boundaries; within it, prediction is practiced and state evolution is guided rather than merely forecast. Steerable, in this program's vocabulary, means exactly this: a model equipped with a harness. The four components map onto the checkpoints of Section 3: constraints onto the declared scaffold behind CP1–CP2, intervention semantics onto CP3, objectives onto the declared expected response of each flywheel cycle, and boundaries onto CP5 together with the predictability research of Section 4. The shepherd principle carries the same idea into practice: a wise shepherd does not map the entire grassland before moving the flock, but studies where the water and grass are (capabilities to maintain), where the wolves and the desert are (the unsupported region), and plans a declared path with supply waypoints (the flywheel). Together: guiding state evolution, not merely forecasting it.

The relation to the field's definition is deliberate. The *Cell* Perspective fixes what to compute: state, action, and time combined into a next-state distribution [15]. CP1-CP4 are that definition made auditable; CP5 adds what it omits. The Perspective's four validation tests all measure how well a model predicts; none measures whether the region of predictability itself has been mapped. In one line: the definition fixes what to compute; predictability research fixes where computing is supported; and steerability keeps it correctable.

The inversion has roots in four independent traditions. Decision theory measures the value of a prediction by the choice it improves [17]. Cybernetics descends from *kybernetes*, the helmsman whose first task is choosing a heading; control has always asked which degrees of freedom to act on [18]. Reinforcement learning introduced world models to serve action, not to score forecasts [5,6].

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 September 2026 doi:10.20944/preprints202609.2101.v1

Oncology long ago separated predictive biomarkers - who responds to this intervention - from prognostic ones, granting selection knowledge a status of its own [19]. Four traditions, one claim: prediction is the means, selection is the end.

3.3. Choosing the Knowledge Map

The second design decision is where the model lives: in what coordinate system are states represented and transitions learned? The proposed answer is a research trajectory (Fig. 3). Network medicine begins from the protein-interaction map - its standard substrate, where nodes are proteins and edges are physical interactions [20]. The route proposed here moves to a dictionary of named gene modules in eleven annotated families (aging hallmarks, organ systems, metabolism, immunity, nutrient response, and related processes), distilled from a published 3,000-pathway methylation atlas [21]. The move is deliberate: a PPI map encodes physical adjacency but no health semantics, while a module map encodes capabilities an intervention can be expected to move. The published benchmark already adopts this dictionary as its state space; the individual-scale and application studies will follow on the same map.

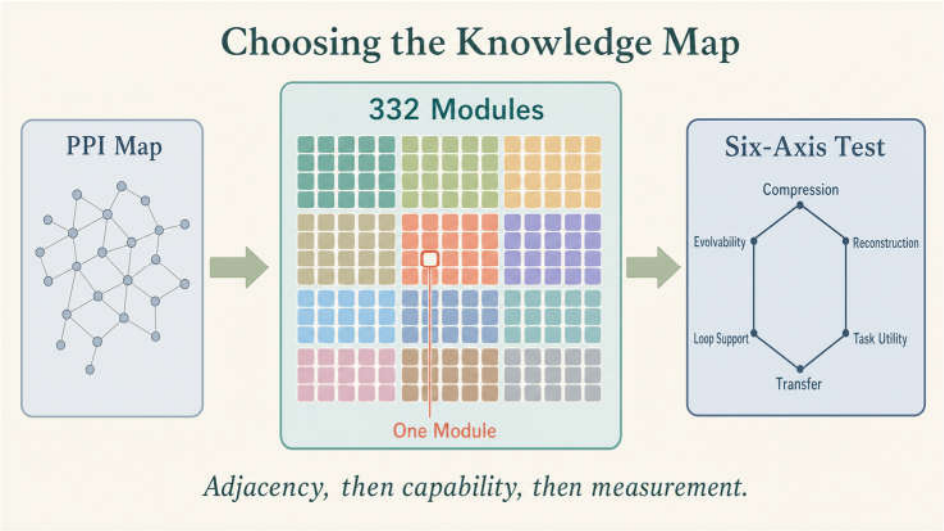


Figure 3. Exploring a better knowledge map. The program's state-representation trajectory. Left: the protein-interaction map - network medicine's standard substrate: nodes are proteins, edges are physical interactions; rich in adjacency, silent on health semantics [20,22]. Middle: the 332-module dictionary - named gene modules in eleven annotated families (aging hallmarks, organ systems, metabolism, immunity, nutrient response, and related processes), adopted as the state space by the published benchmark [23], with the individual-scale and application studies to follow on the same map; a capability an intervention can be expected to move. Right: the six-axis evaluation protocol that judges any candidate coordinate system - compression, reconstruction, task utility, transfer, intervention-loop support, evolvability - turning map choice from taste into measurement. Message: which knowledge map to use is itself a measurable, task-dependent research question.

A compression claim needs an evaluation, not an assertion. Six axes define what such an evaluation must measure: compression, reconstruction, task utility, transfer, intervention-loop support, and evolvability. Which knowledge map to use is thus itself a measurable, task-dependent question, not a matter of taste. One boundary deserves explicit statement: CP4 makes no identifiability claim in the causal-inference sense. Transitions are calibrated predictions under declared assumptions; Section 4 states how the agenda will probe that boundary.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 September 2026 doi:10.20944/preprints202609.2101.v1

3.4. Steerable State Transitions, Defined

The title term deserves its own definition. *Steerable state transitions* are constrained moves within a mapped region. The state representation is chosen by measurement, not convention. The expected response is declared before acting. At each step, only a few moves remain supported, by predictability research. Deviations are inspected at named checkpoints, not absorbed as noise.

This definition has a spatial reading. A temporal transition can be transposed onto the dependency graph: choosing which upstream module to adjust, with measured confidence in the modules it steers (Fig. 1). When such edges hold - aging hallmarks associated with organ aging, organ aging associated with its phenotypes - a candidate upstream capability generates a directional hypothesis about what lies downstream. The temporal question, what comes next, becomes a spatial one: which node to choose. A first such map is under construction and will be reported separately.

Spatial structure does not answer the temporal question by itself. It organizes the question. A dependency graph specifies which transition hypotheses are worth testing and which variables should be measured together; paired observations then decide what survives. Structure constrains, and time tests.

The two routes also differ in price. The default route waits for large longitudinal cohorts: tens of thousands of persons, years of follow-up, national-program budgets. The spatial route runs on what already exists. Cross-sectional resources carry the structure; small paired samples from public cohorts test it; each measurement cycle takes months, with no new enrollment. The exchange rate is crude but real: structure bought with space repays time. This is a hypothesis about research efficiency, not a claim that cross-sectional structure substitutes for prospective longitudinal validation.

Four features follow: bounded search instead of open-ended optimization; evidence graded to its tier; negative results kept as maps of where practice is not yet supported; and each care cycle doubling as data. Accumulated over cycles, such transitions sketch what steerable medicine would mean - a practice ideal, not a delivered system. Section 5 states what the route must still demonstrate.

4. The Agenda, Question by Question

Two published systems anchor the agenda (Table 2). SteeraMed [22] implements N-of-1 reasoning on the module space; its companion benchmark [23] turns the same state space into a public resource - a 332 x 1,916 drug-module matrix with formalized state-transition rules. On these foundations, the agenda schedules four questions.

Can clocks read intervention response? Paired before-and-after observations - the data type the *Cell* Perspective names as the field's binding constraint - will be read through two lenses: a single-score clock and the module map. The contrast is the finding, whichever way it falls.

Do cellular priors reach patients? A conditional-transfer study with pre-registered commitments will test cell-to-patient transfer module family by module family. The immune axis is where transfer is most plausible, and the negative regions are expected to be the map's most useful output.

Does medical knowledge organize measurably? A dependency map of intrinsic capabilities, hallmark to organ, is under construction. It will be scored by declared statistical defenses and by independent-cohort direction replication before any structural claim is made.

Can concurrent interventions be told apart? Decomposability of multi-intervention responses is the prospective frontier. A concept framework exists; the empirical work is scheduled.

The agenda states its questions before it knows the answers; each study will report its own quantitative details. Negative results are not failures of the program - they are the program's maps of where prediction is not supported.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 September 2026 doi:10.20944/preprints202609.2101.v1

5. Positioning and Verdict

5.1. What Is Different Here

Three elements distinguish this route. First, the agenda is built on the data type the *Cell* Perspective names as the binding constraint: individual-scale paired intervention observations. Second, an explicit predictability layer: transferability and response-density are treated as measurable properties of the world, not as background assumptions. Third, every claim carries a stated test and boundary (Table 2) - no other proposal we are aware of carries a ledger of how its own claims could fail.

Table 2. Claims, planned tests, and boundaries.

Module	Claim	Planned test	Status	Boundary
Theory	A life world model needs five auditable checkpoints, and its state can be a computed mIC profile	Prototype + worked deviation-inspection case [7]; six-axis evaluation of the coordinate system	Anchor published [7]; evaluation scheduled	Not a clinical decision system; case study is a thought experiment
Systems	Module state space adds measurable value to network-medicine discovery and supports N-of-1 ranking	Baseline-controlled repositioning enrichment and candidate recovery; 332 × 1,916 public benchmark [23]	Anchor published [22,23]	Not claimed above public baselines on continuous metrics; prospective N-of-1 pending
Individual scale	Whether clocks or module maps read intervention response, and whether cellular priors reach patients, are measurable questions	Paired before-and-after contrast, clock vs module map; pre-registered conditional-transfer test	Scheduled	No individual-scale claims made here; boundaries will be reported by those studies
Application	Whether medical knowledge organizes measurably, hallmark to organ, is a testable claim	Dependency map scored by declared statistical defenses + independent-cohort direction replication	Under construction	Auditable hypothesis generation, not established causality
Future	Whether concurrent-intervention effects can be told apart is itself a measurable question	Concept framework exists; decomposability study scheduled	Prospective	No empirical claims made here

5.2. The Verdict

The route can be read in three questions. *Predict* - answering "what is": whether module maps read per-person intervention response while single-score clocks act as drift detectors is the first scheduled question. *Simulate* - answering "what would happen under intervention": single-step

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 September 2026 doi:10.20944/preprints202609.2101.v1

transition rules and the 332 x 1,916 benchmark are in hand; multi-step rollout is not yet scheduled. *Steer* - answering "why did the prediction fail and what to revise": the CP5 five-gate interface and the data flywheel are the working answer, with prospective validation pending. The graded verdict: an auditable framework for studying predictability and revising state-transition hypotheses. Whether this framework delivers prospective steering utility remains an open empirical question. Steerability is drawn as a property of the model (equipped with a harness; its failures diagnosable and its revisions costed), not of biological systems.

6. The Road Ahead

The checkpoints close into a flywheel: measure state, choose an intervention with a declared expected response, intervene, re-measure, inspect deviation at the failing checkpoint, and revise model or plan. Each cycle can generate both decision-relevant evidence and model-updating data, subject to appropriate governance [7].

How might a spatial structure actually encode temporal ability? Several routes are open, and the published anchors make each concrete; they have kin in structure-informed machine learning, and the open question is which survive contact with biological data. (i) *Priors*: the dependency graph enters as a prior or regularizer, so a predicted response stays near its structural neighborhood. (ii) *Propagation*: message-passing models run transitions along supported edges rather than free vectors. (iii) *Masks*: predictability research yields an evidence-supported subgraph; prediction runs only there, and the search collapses by construction. (iv) *Hierarchy*: root-to-organ-to-phenotype layers give state-space models a causal reading of depth. (v) *Driver selection*: control theory picks the fewest nodes that steer the rest - the steering points of Section 3.4. (vi) *Structure learning in the loop*: each before-and-after pair updates edge confidence, so the map co-evolves with practice. (vii) *Multi-fidelity data*: cross-sectional structure is abundant, paired longitudinal data are scarce; a good encoding lets the former carry the latter. In short, structure constrains the search, and time refines the structure. A first empirical structure for these routes - a dependency map of hallmark-to-organ edges - is under construction.

Five questions remain open, each with a defined attack path. Multi-step simulation needs longitudinal cohorts with repeated interventions. Cross-scale coupling needs joint cell-plus-patient measurement. Absolute mIC calibration must hold across platforms and populations. Prospective N-of-1 validation is pending [22]. And decomposability of concurrent-intervention effects is the route's prospective frontier. The predictability-first ordering also aligns with how interventions are actually regulated: review bodies ask for mechanistic rationale, not accuracy alone. A system that states which modules an intervention touches, and where its predictions are not supported, produces the mechanism-clear, boundary-explicit reasoning such review demands.

The route is falsifiable by design. It would be supported on three conditions. Structure-constrained models should beat unconstrained baselines on held-out interventions. Predictability maps should identify low-transfer settings in advance. And structural priors should improve calibration, transfer, or sample efficiency, not training fit alone. It would require revision under the mirror conditions: no held-out gain from structural constraints, predictability maps failing in new cohorts, or module maps offering no advantage over reasonable random compressions across the full task battery. Each condition names a measurement the flywheel can run - the agenda states, in advance, what would count against it.

7. Boundaries and Conclusion

This perspective describes a framework under construction, not a validated clinical system. The framework is deductively constrained but its transitions are calibrated predictions, not identified causes. The individual-scale questions are scheduled, not settled: their quantitative answers belong to the studies now underway and will be reported as they appear. No clinical decision capability is claimed.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 September 2026 doi:10.20944/preprints202609.2101.v1

An intervention-oriented biomedical world model is not valuable merely because it predicts more accurately. Its value lies in whether it improves auditable selection under uncertainty. It is a harness for biological state evolution. The practical route begins not with predicting every future state, but with learning - explicitly and empirically - where prediction is possible, where it fails, and what measurement would move that boundary. The scaffold is declared; the map turns temporal questions into spatial ones; predictability research bounds where to act; five gates correct the system when reality disagrees. The route in one line: the problem is not prediction but selection - prediction is the means, selection the end; structure constrains, and time tests; failed transfer is a result. The next decade belongs to models that know what they cannot yet see, and that knowledge is engineered, not assumed.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The published studies cited as [7,22,23] carry their own DOIs, under which data availability is described. The SteeraMed platform is at https://steeramed.com.

Conflicts of Interest: J.X. is the founder of DeepoMe Limited (Beijing, China), which develops the SteeraMed platform discussed in this Perspective. The author has a potential commercial interest in the concepts and tools described herein.

References

1. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biology*. 2013;14(10):R115. doi:10.1186/gb-2013-14-10-r115.
2. Fuentealba M, Rouch L, Guyonnet S, Lemaitre JM, de Souto Barreto P, Vellas B, et al. A blood-based epigenetic clock for intrinsic capacity predicts mortality and is associated with clinical, immunological and lifestyle factors. *Nature Aging*. 2025;5:1207–1216. doi:10.1038/s43587-025-00883-5.
3. Fard P, Azhir A, Rezaii N, Tian J, Estiri H. An N-of-1 Artificial Intelligence Ecosystem for Precision Medicine. *arXiv preprint*. 2025. arXiv:2510.24359.
4. Pearl J. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press; 2009.
5. Hafner D, Pasukonis J, Ba J, Lillicrap TP. Mastering diverse control tasks through world models. *Nature*. 2025;640(8059):647–653. doi:10.1038/s41586-025-08744-2.
6. Ha D, Schmidhuber J. World Models. *arXiv preprint*. 2018. arXiv:1803.10122.
7. Xiong J. *World Models for Biomedicine: A Steerability Framework. Preprints*. 2026. doi:10.20944/preprints202605.0366.v1.
8. Qazi MA, Nadeem M, Yaqub M. Beyond Generative AI: World Models for Clinical Prediction, Counterfactuals, and Planning. *arXiv preprint*. 2025. arXiv:2511.16333.
9. Saeed N, Hassan S, Khan S, Qazi MA, Maier-Hein KH, Khan S, Yaqub M. Medical World Model: From Passive Prediction to Active Simulation in Medicine. *Preprints*. 2026;2026042168. doi:10.20944/preprints202604.2168.v1.
10. Yang Y, Wang ZY, Liu Q, Sun S, Wang K, Chellappa R, Zhou Z, Yuille A, Zhu L, Zhang YD, Chen J. Medical World Model: Generative Simulation of Tumor Evolution for Treatment Planning. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2025:8319–8329. arXiv:2506.02327.
11. Wei Z, Ma R, Wang Z, Li Z, Song S, Zheng S. VCWorld: A Biological World Model for Virtual Cell Simulation. *International Conference on Learning Representations*. 2026. arXiv:2512.00306.
12. Callahan TJ, Beckwith Z, Merth T, van der Poel C, Lewis A, Lemos P. Virtual Cells as Causal World Models: A Perspective on Evaluation. *NeurIPS 2025 AI4D3 Workshop*. 2025.
13. Business Wire. *GSK Licenses Noetik's AI Foundation Models in Anchor Partnership to Transform Cancer Therapeutic Research and Development*. January 8, 2026.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 September 2026 doi:10.20944/preprints202609.2101.v1

14. Zhang HG, et al. The virtual biotech: a multi-agent AI framework for therapeutic discovery and development. *Science*. 2026. doi:10.1126/science.aeg6779.
15. Noori A, Fishman N, Fang A, Fesser L, Zitnik M. World models for biomedicine. *Cell*. 2026;189(19):5845–5869. doi:10.1016/j.cell.2026.08.032. PMID: 42753693.
16. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*. 2022;35:27730–27744. arXiv:2203.02155.
17. Howard RA. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*. 1966;2(1):22–26. doi:10.1109/TSSC.1966.300074.
18. Wiener N. *Cybernetics: Or Control and Communication in the Animal and the Machine.* Cambridge, MA: MIT Press; 1948.
19. Sawyers CL. The cancer biomarker problem. *Nature*. 2008;452(7187):548–552. doi:10.1038/nature06913.
20. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: A network-based approach to human disease. *Nature Reviews Genetics*. 2011;12(1):56–68. doi:10.1038/nrg2918.
21. Xiong J. Next Generation Aging Clock: A Novel Approach to Decoding Human Aging Through Over 3000 Cellular Pathways. *bioRxiv*. 2024. doi:10.1101/2024.06.18.599632.
22. Xiong J. *SteeraMed: A Biomedical World Model for N-of-1 Intervention Reasoning Across Chronic Diseases and Aging. Preprints*. 2026. doi:10.20944/preprints202605.1578.v1.
23. Xiong J. Toward a self-learning AI agent for drug repurposing: Building human-scale representations for virtual patients. *Preprints*. 2026. doi:10.20944/preprints202608.0998.v1.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.