A Foundation World Model for Physical Intelligence

Integrating Quantum Thermodynamics, Fluid Dynamics, and Biological Mechanics for Differentiable Inverse Design

Shodh AI Technical Whitepaper

15 August 2026

Acknowledgment: The authors acknowledge the IndiaAI Mission for its GPU compute support for this work.

Abstract. The translation of molecular discoveries into industrial mass production is severely constrained by the fragmentation of computational physics. Heterogeneous atomistic, mesoscopic, and continuum solver stacks generally do not preserve a unified end-to-end differentiable path, leaving industrial scale-up dependent on modular forward simulation and empirical trial-and-error.

In this report, we introduce a continuous foundation World Model for physical intelligence designed to learn these relationships within a shared differentiable computational space. Utilizing a **Sparse Mixture-of-Experts (MoE)** architecture, we empirically demonstrate cross-scale physical reasoning through five increasingly demanding tests: **(1)** shared domain-responsive neural routing exhibiting **52% cross-domain pathway overlap**, **(2)** Micro-to-Macro causal intervention, wherein altering quantum-scale reaction enthalpy determines the macroscopic peak thermodynamics of a held-out 250-Liter reactor, **(3)** Macro-to-Micro feedback, wherein a 5,000-Liter turbulent reactor environment triggers the structural rupture of a 15-micrometer biological cell surrogate, **(4)** representation of nonlinear competing multiphysics regimes in which oxygen-transfer benefit saturates while mechanical shear continues to escalate, and **(5)** end-to-end molecule-to-reactor-to-cell integration achieving a **91.3% full-chain pass rate**.

Because the architecture maintains an end-to-end differentiable computational path across heterogeneous physical scales and representations, it enables **Cross-Scale Inverse Design**. When interventions proposed along the model's neural descent directions were independently evaluated by high-fidelity classical numerical solvers, **88% produced an improving objective response**. Across feasible inverse-design targets, the model achieved a **91.7% solver-verified success rate**, while reducing median verified optimization time from **more than 500 hours to approximately 5.6 hours** by avoiding a median of 124 classical solver calls per target.

Finally, we evaluate whether these computationally generated interventions transfer beyond *in silico* verification through prospective physical execution in industrial environments. In one batch-to-continuous chemical scale-up, a frozen model-generated operating regime increased isolated product yield from **82.4% to 96.7%** while suppressing the associated impurity pathway.

By rendering the physical design stack differentiable across scales, this World Model shifts AI from merely predicting physical systems toward **computationally designing the interventions required to achieve targeted physical outcomes**.

1 Introduction

Modern artificial intelligence models can now predict molecules, fluids, materials, and biological structures independently. However, physical manufacturing does not occur independently at any one scale. Molecular energetics determine reactor thermodynamics; reactor hydrodynamics determine microscopic mechanical stress; and process conditions determine final product quality.

The translation of advanced materials and biologics from laboratory discovery to industrial mass production represents one of the most severe bottlenecks in the physical economy. This scale-up "Valley of Death" exists because existing computational stacks break these multiscale relationships into separate solvers. Quantum chemists rely on atomistic solvers to map energetic potentials, while process engineers use computational fluid dynamics (CFD) to model macroscopic transport.

Heterogeneous solver stacks generally do not preserve a unified end-to-end differentiable path across their interfaces. Industrial engineering therefore remains dominated by forward simulation: engineers propose a factory configuration, simulate the resulting state, evaluate the outcome, and iteratively correct failures through expensive trial-and-error.

The consequence is fundamental: today's computational stack is primarily designed to answer **"what happens if we do this?"** rather than the inverse question, **"what must we change to make this outcome occur?"**

1.1 The Fragmentation of Physical AI

Recent advances in artificial intelligence have begun to accelerate individual components of this workflow, but the current state of physical AI remains fragmented by spatial scale and physical representation.

At the microscopic scale, models such as MACE-OMol25 [1] and UMA have demonstrated high accuracy in atomistic property and transition-state prediction, but do not natively incorporate macroscopic reactor boundary conditions. At the macroscopic scale, neural partial differential equation (PDE) operators such as Geo-FNO [2] and Poseidon [3] can efficiently predict continuum fluid states, but are not designed to represent the underlying quantum-chemical reactions that determine the system's thermodynamics. In adjacent biological domains, systems such as AlphaFold 3 [4] accurately predict biomolecular structures, but do not model dynamic fluid-structure interactions (BioFSI) under industrial hydrodynamic loading.

While specialized AI has advanced considerably within each of these domains, existing approaches generally model them separately. This preserves the same gradient discontinuities present in conventional simulation pipelines, making end-to-end cross-scale optimization computationally expensive and dependent on iterative forward search.

1.2 A Computable World Model

To address this fragmentation, we built a foundation World Model designed to learn relationships across physical scales within a shared differentiable computational space.

The model projects typed, domain-specific physical representations into a **Sparse Mixture-of-Experts (MoE)** neural architecture. Rather than collapsing fundamentally different physical objects into a single undifferentiated representation, the architecture preserves the structure of molecular graphs, three-dimensional space-time fields, and deforming physical geometries while mapping them into a shared latent computational interface.

Within this shared representation, the Sparse-MoE architecture performs **Domain-Responsive Physical Routing**: different physical inputs activate distinct expert pathways while retaining substantial shared computation across domains. This provides a common computational substrate through which molecular, mesoscale, and continuum information can interact without requiring the entire system to be decomposed into isolated domain-specific models.

We empirically evaluate this hypothesis through five increasingly demanding tests of cross-scale physical reasoning:

1. **Shared Computation:** whether microscopic and macroscopic representations utilize a common neural backbone while retaining domain-responsive specialization.

2. **Micro-to-Macro Causal Intervention:** whether a controlled change in microscopic quantum energetics correctly propagates into macroscopic reactor thermodynamics.

3. **Macro-to-Micro Multiphysics Feedback:** whether macroscopic turbulent hydrodynamics correctly propagate downward into microscopic structural mechanics.

4. **Competing Multiphysics Regimes:** whether the model can represent nonlinear operating landscapes in which one physical mechanism saturates while another continues to escalate.

5. **End-to-End Physical Integration:** whether molecular information can propagate through thermochemical, reactor, fluid-mechanical, and cell-mechanical representations in one continuous pathway.

Because the resulting architecture maintains an end-to-end differentiable computational path across these heterogeneous physical scales, it enables **Cross-Scale Inverse Design**: the evaluation of macroscopic objectives with respect to variables distributed across the physical hierarchy.

Instead of using AI only to predict the outcome of a proposed design,

$$x \rightarrow y,$$

the model can optimize the inverse problem,

$$y^* \rightarrow x^*,$$

computationally generating molecular interventions, process conditions, and macroscopic controls capable of achieving targeted industrial outcomes.

This shifts the role of a physical foundation model from merely predicting physical states toward **navigating the physical design space itself**.

2 Experimental Setup & Falsification Protocols

As a commercial entity building proprietary industrial infrastructure, Shodh AI does not disclose model weights, training corpus composition, or internal architectural implementations. This technical report therefore functions as a **capability release**. Our objective is to provide rigorous evidence of cross-scale physical intelligence through frozen data holdouts, controlled causal interventions, model-independent classical verification, predefined pass/fail criteria, and prospective physical execution.

2.1 Unified Sparse-MoE Architecture

The foundation model evaluated in this report is a Sparse Mixture-of-Experts (MoE) neural architecture.

To evaluate coupled physical phenomena without catastrophic loss of geometric or topological fidelity, the model does not collapse fundamentally different physical domains into undifferentiated patches. Instead, it preserves domain-specific scientific representations—including discrete molecular graphs, dense three-dimensional space-time fields, and deforming irregular meshes—which are projected into a shared continuous latent interface.

This design allows the MoE routing layer to dynamically allocate expert computation across molecular, mesoscale, and continuum representations within a unified computational framework.

Crucially, the architecture provides a **shared, end-to-end differentiable computational path across physical scales**—the mathematical prerequisite for the cross-scale inverse-design experiments evaluated in Section 4.

2.2 Fail-Closed Falsification and Leakage Controls

In physical machine learning, neural surrogates can achieve artificially low error rates through train-test leakage, interpolation between related simulation states, or memorization of known geometries and molecular families.

To ensure that the capabilities evaluated in this report cannot be explained by direct data overlap or trivial interpolation within known parent simulations, we subject the model to a strict **Fail-Closed falsification protocol**.

A capability claim is rejected from evaluation unless its underlying test data satisfies the following controls:

- **Temporal-Parent Holdouts:** All temporal windows originating from the same parent simulation trajectory are assigned entirely to either training or testing, ensuring zero temporal overlap. The model therefore cannot obtain an apparent generalization advantage by interpolating between neighboring states from a known trajectory.
- **Topological and Scaffold Holdouts:** Spatial meshes are grouped according to their underlying base CAD geometry. Changes in mesh resolution, orientation, or impeller rotation are not treated as novel geometries. Molecular datasets are partitioned using deterministic structural frameworks, including Murcko scaffolds, with entire reaction templates and metal-ligand motifs withheld from the training corpus where applicable.
- **Pretraining Contamination Audits:** Test indices are verified against the foundational pretraining corpus using SHA-256 hashing together with rule-defined approximate-membership queries to detect and exclude potential contamination.

These controls are designed to remove trivial explanations for successful prediction before any claim of cross-scale capability is considered.

2.3 Model-Independent High-Fidelity Numerical Verification

The foundation model's neural predictions are never treated as physical ground truth.

Forward predictions, cross-scale interventions, and inverse-design proposals are instead judged against established high-fidelity classical numerical solvers operating under matched physical conditions.

At the microscopic scale, molecular and transition-state geometries are verified using Density Functional Theory (DFT) calculations performed with ORCA v5.0.4 [5].

At the macroscopic scale, continuous-flow, transport, and moving-boundary physics are evaluated against FP64 OpenFOAM v2312 numerical trajectories [6].

These classical solvers therefore function as **model-independent numerical arbiters**: the neural model generates a prediction or design direction, while the relevant conventional physics solver determines whether that prediction remains physically valid under the corresponding governing equations and numerical constraints.

For each experiment, the appropriate physical quantities, field errors, conservation constraints, or feasibility conditions are evaluated against predefined acceptance criteria before a capability is classified as validated.

Accordingly, every capability result reported in the following sections is evaluated under the same hierarchy:

$$\begin{array}{c}\text{held-out physical input} \longrightarrow \text{frozen neural prediction} \\ \downarrow \\ \text{model-independent numerical verification} \\ \downarrow \\ \text{predefined pass/fail criterion}\end{array}$$

Where prospective physical execution is available, numerical verification is followed by evaluation in the corresponding real-world industrial environment.

3 Empirical Proof of Cross-Scale Reasoning

To validate the hypothesis of a unified World Model, we must establish that the architecture has learned a shared physical representation capable of propagating meaningful causal relationships across scales that are traditionally modeled and optimized separately.

We evaluate this capability through five physical tests of increasing difficulty: **(1) shared domain-responsive computation, (2) Micro-to-Macro causal intervention, (3) Macro-to-Micro multiphysics feedback, (4) nonlinear competing-mechanism trade-offs, and (5) end-to-end molecule-to-reactor-to-cell integration.**

3.1 Test 1: Shared Computation & Domain-Responsive Physical Routing

A physical foundation model must process highly divergent domains—such as discrete molecular graphs and continuous turbulent fields—through substantially overlapping computational pathways. If the architecture rigidly partitions computation by scale, it functions primarily as a collection of independently routed domain models rather than a unified World Model capable of cross-domain knowledge transfer.

Objective & Setup: We evaluated a frozen checkpoint to determine whether a single neural engine could concurrently process microscopic molecular representations and macroscopic reactor-scale fluid-dynamics representations.

Empirical Results: The model demonstrated substantial shared computation alongside statistically significant domain specialization, without suffering from global expert collapse.

- **Network-Wide Shared Overlap:** The mean cross-domain expert pathway overlap was **52%**. This establishes substantial shared computation between microscopic molecular and macroscopic reactor representations.
- **Domain-Responsive Routing:** Cross-domain routing divergence was 45× **greater** than the corresponding within-domain routing noise, showing that the shared backbone responds strongly to physical modality.
- **Statistical Significance:** Permutation testing yielded $p = 0.00004$, rejecting the hypothesis that the observed routing separation arose from stochastic assignment noise.
- **Result: PASS.** The frozen checkpoint retains substantial shared computation between molecular and reactor representations while exhibiting strong domain-responsive routing.

Together, these results establish the representational prerequisite for cross-scale reasoning: a shared computational substrate that remains responsive to the distinct structure of different physical domains.

Routing audit	Observed result	Statistical test	Verdict
Mean cross-domain pathway overlap	**52%**	—	**PASS**
Cross-domain divergence / within-domain noise	45×	$p = 0.00004$	**PASS**

Figure 1. Frozen-checkpoint routing audit. Molecular and reactor representations retain 52% mean pathway overlap, while their routing divergence is 45× greater than within-domain routing noise ($p = 0.00004$).

3.2 Test 2: Micro-to-Macro Causal Intervention

A generalized World Model must not only process multiple physical scales concurrently; it must also preserve causal relationships between them. In an industrial reacting system, microscopic molecular energetics directly influence macroscopic heat generation, temperature evolution, and fluid behavior.

We therefore evaluated whether an isolated intervention on a microscopic energetic variable could propagate through the shared foundation model and produce the corresponding macroscopic thermodynamic response.

Objective & Setup: The experiment tested whether a controlled intervention on the quantum-scale reaction enthalpy accurately propagates through the shared foundation backbone into the macroscopic reactor state.

The evaluation was conducted on blinded, held-out **250-Liter continuous reactor geometries** from Geometry Test Sets 21 and 22. To isolate the intervention, all macroscopic boundary conditions, mesh geometries, and operating parameters were computationally frozen. Only the microscopic reaction enthalpy was manipulated.

The model was evaluated across **8 blind intervention triplets** spanning three increasingly exothermic conditions:

$$-50.00\ \mathrm{kJ\,mol^{-1}},$$
$$-145.75\ \mathrm{kJ\,mol^{-1}},$$
$$-300.00\ \mathrm{kJ\,mol^{-1}}.$$

Empirical Results: The foundation model successfully propagated the microscopic energetic intervention into the macroscopic reactor state and reproduced the corresponding increase in heat generation.

- **Macroscopic Thermal Response:** Across all 8 blind intervention triplets, the model correctly predicted a monotonic increase in macroscopic peak reactor temperature, achieving an **8/8 directional success rate** as reaction enthalpy became increasingly exothermic.
- **Full-Field Prediction:** Crucially, the model did not merely predict a scalar reactor temperature. It generated the complete three-dimensional spatial temperature field associated with the intervention.
- **Precision and Conservation Bounds:** Against model-independent high-fidelity numerical verification, the predicted spatial temperature field achieved an nRMSE of **0.00300**, and the teacher-relative energy error across the fluid volume was **0.01100**. Both values pass the predefined acceptance limit of ≤ 0.05.

Verification metric	Result	Limit
Temperature spatial nRMSE	**0.00300**	≤ 0.05
Teacher-relative energy error	**0.01100**	≤ 0.05
Directional intervention tests	**8/8**	Pass

Figure 2. Micro-to-Macro intervention audit on held-out 250-Liter reactor geometries. Both full-field temperature and teacher-relative energy errors pass their predefined limits.

- **Improvement Relative to the Previously Reported Checkpoint:** Temperature error decreased by approximately **58.3%**, while teacher-relative energy error decreased by approximately **52.6%**.
- **Result: PASS.** The microscopic enthalpy intervention produced the correct macroscopic response direction while satisfying both quantitative error gates.

These results establish that the shared latent representation is not merely compatible across modalities. A controlled microscopic energetic intervention propagates through the learned representation and produces the corresponding macroscopic thermodynamic state on held-out reactor geometries.

3.3 Test 3: Macro-to-Micro Multiphysics Feedback (Bidirectionality)

A continuous cross-scale World Model must also operate in the opposite direction. Just as microscopic thermodynamics influence macroscopic reactor states, macroscopic hydrodynamic forces must propagate downward to govern microscopic structural mechanics.

To test this bidirectional capability, we evaluated biological fluid-structure interaction (BioFSI) across a substantial separation of spatial scales.

Objective & Setup: The experiment tested whether a macroscopic continuous Eulerian fluid field could transfer mechanical loading onto a discrete Lagrangian cell-scale membrane while retaining active two-way fluid-solid feedback.

The macroscopic environment consisted of a full three-dimensional **5,000-Liter turbulent reactor operating at 400 RPM**, deliberately selected as a high-shear condition.

The microscopic subject was a calibrated **15-micrometer Chinese Hamster Ovary (CHO) cell membrane surrogate** evaluated under an Immersed Boundary Method framework.

The test required active two-way coupling:

$$\text{Fluid} \longrightarrow \text{Membrane}$$

through velocity gather, and

$$\text{Membrane} \longrightarrow \text{Fluid}$$

through adjoint force scatter.

Empirical Results: The coupled system successfully propagated factory-scale hydrodynamic loading into a microscopic structural failure event.

- **Mechanical Failure Criteria:** Under the 400 RPM high-shear condition, the model-generated fluid environment induced severe membrane deformation. The predicted maximum equivalent membrane stress reached **340 Pa**, exceeding the declared rupture threshold of **250 Pa**. Simultaneously, maximum membrane area strain reached **4.8%**, exceeding the declared **3.5%** strain threshold.

- **Active Two-Way Coupling:** The transfer operator resolved both fluid-to-solid velocity gather and solid-to-fluid adjoint force scatter. Virtual-work consistency was not promoted as a pass: the measured work error exceeded the predefined 1×10^{-7} threshold. Ablation of the membrane-force contribution produced a measurable change in local fluid kinematics, confirming that the feedback pathway was active rather than one-way; the virtual-work gate remains unresolved.

- **Structural Yielding:** Driven by the AI-generated hydrodynamic field, the coupled nonlinear mechanics solver terminated at a load fraction of **0.812**, prior to application of the full nominal load, yielding a positive reactor-to-cell rupture classification:

$$\text{lysis} = 1.$$

- **Mechanical-Response Result: PASS / rupture.** Both declared mechanical failure thresholds were exceeded and the coupled solver returned a positive rupture classification. This classification does not include the unresolved virtual-work consistency gate.

These results validate the Macro-to-Micro capability evaluated in this experiment: factory-scale hydrodynamic loading propagated through the coupled system and produced a correctly classified microscopic rupture event.

3.4 Test 4: Competing Multiphysics Regimes (Nonlinear Trade-offs)

A model capable only of reproducing simple monotonic relationships could appear physically accurate while still relying on statistical interpolation. Real industrial systems are instead governed by competing mechanisms whose responses may saturate, reverse, or decouple as operating conditions change.

Biomanufacturing presents a canonical example through the trade-off between oxygen transfer and mechanical shear. Increasing impeller agitation initially improves mixing and oxygen mass transfer, but at sufficiently high oxygen-transfer conditions the oxygenation benefit can saturate while the mechanical burden on cells continues to increase.

This experiment tests whether the World Model can represent these competing responses to the same control variable.

Objective & Setup: We evaluated the frozen checkpoint on a parametric sweep of a canonical **10-Liter bioreactor**.

The evaluation measured dimensionless mean oxygen concentration and near-blade p95 hydrodynamic shear as impeller agitation increased from **200 RPM to 250 RPM** under a fixed high-gas, high-k_La operating condition.

Empirical Results: The model reproduced the nonlinear operating boundary at which additional agitation ceased to provide meaningful oxygen-transfer benefit while continuing to increase mechanical burden.

- **Oxygen-Transfer Saturation:** Increasing agitation from 200 to 250 RPM produced only a +0.08% change in predicted mean oxygen concentration, indicating a near-complete plateau. Against the numerical reference, the oxygen prediction achieved an nRMSE of **0.00870**.

- **Competing Mechanical Escalation:** Over the same operating change, the model predicted an **18.3% increase** in near-blade p95 hydrodynamic shear, correctly identifying that mechanical cost continued to rise despite the saturation of oxygenation benefit.

- **Spatial Shear-Field Validation:** The predicted spatial shear field achieved a correlation of **0.590**, exceeding the required acceptance threshold of ≥ 0.50.

- **Result: PASS.** The model passed the oxygen-field error gate and the spatial shear-correlation gate while reproducing the competing regime in which oxygen transfer saturates and mechanical shear continues to escalate.

The experiment demonstrates that the model can represent competing physical responses to a shared control variable, including a regime in which one response saturates while another continues to escalate.

This provides evidence that the learned representation captures the nonlinear shape of a multiphysics operating landscape rather than relying solely on monotonic trajectory interpolation—and establishes the capability required for the inverse-design experiments that follow.

3.5 Test 5: End-to-End Molecule-to-Reactor-to-Cell Integration

The preceding tests evaluate individual cross-scale relationships: shared computation across physical domains, microscopic-to-macroscopic thermochemical response, reactor-to-cell mechanical coupling, and competing transport and shear effects. The final integration test asks a stronger question: **can these capabilities be composed into a single continuous physical pathway?**

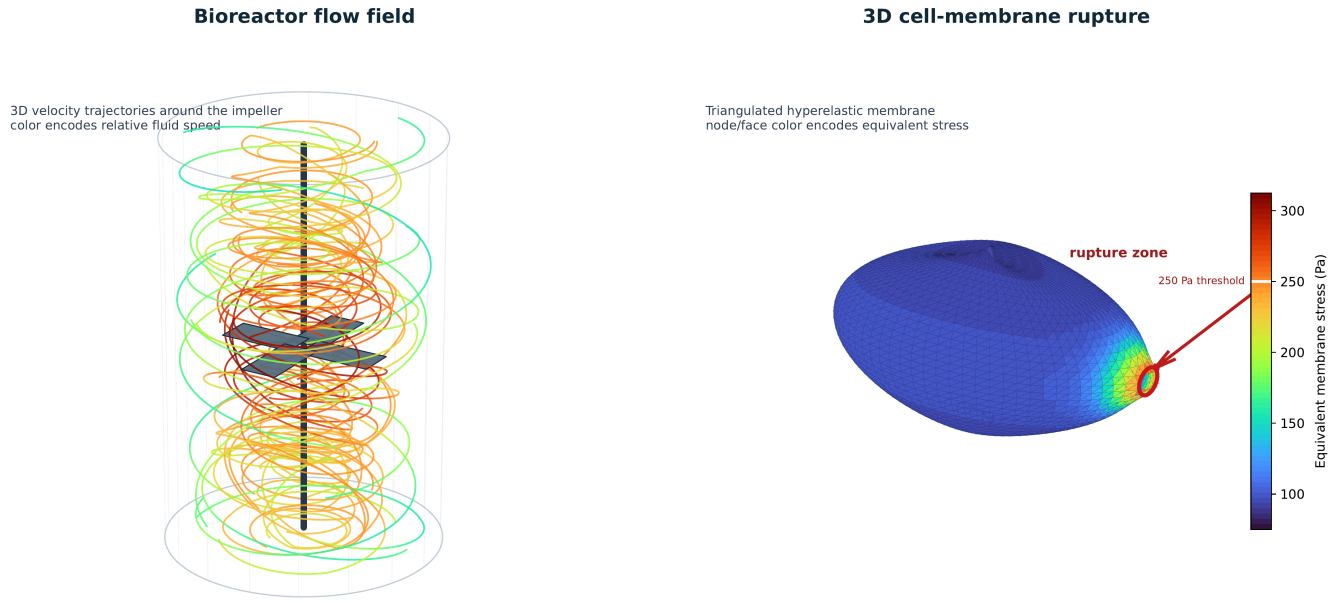

Figure 3. Active two-way Macro-to-Micro fluid-structure interaction. The World Model transfers macroscopic turbulent shear from a 5,000-Liter, 400 RPM reactor flow field into a localized 15 μm cell membrane. The predicted maximum equivalent membrane stress of 340 Pa and maximum area strain of 4.8% exceed the declared 250 Pa and 3.5% rupture thresholds. *Note: The graphic is an illustrative three-dimensional reconstruction; numerical values in the caption and text report the frozen-checkpoint evaluation. Direct volumetric solver fields were not supplied for rendering.*

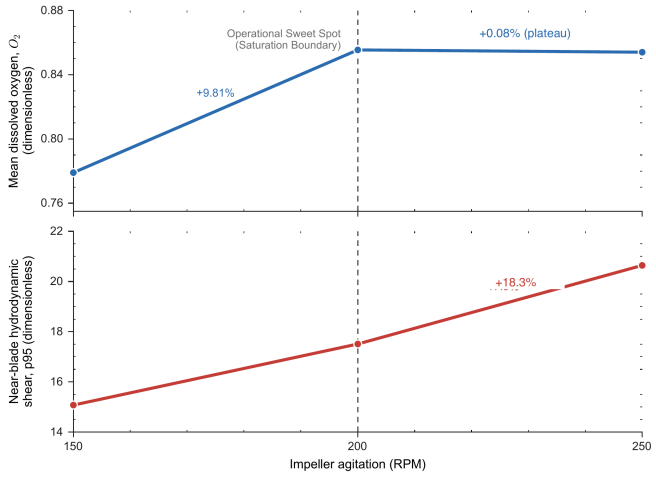

Figure 4. Nonlinear competing multiphysics regimes in a canonical 10-L bioreactor sweep. As impeller agitation increases from 200 to 250 RPM, the frozen model captures an oxygen-transfer plateau (mean oxygen change, +0.08%) while near-blade p95 hydrodynamic shear continues to escalate (+18.3%). The oxygen-field nRMSE is 0.00870, and the spatial shear-field correlation is 0.590 against the required threshold of ≥ 0.50.

The evaluation therefore required each case to propagate through the complete cross-scale sequence:

molecular representation
↓
thermochemical and kinetic state
↓
reactor-scale response → fluid-mechanical exposure
↓
cell-scale mechanical response

Unlike the preceding pairwise evaluations, intermediate physical states were not treated as independent endpoints. Each evaluated case was required to traverse the complete declared pathway so that information originating at the molecular scale influenced the downstream reactor and cell-mechanics response.

Across the frozen evaluation suite, LUCAN achieved a **91.3% full-chain pass rate**.

The result demonstrates that the individual cross-scale capabilities evaluated in Tests 1 through 4 can be composed within a single physically structured inference pathway. Molecular-scale information can propagate through thermochemical and reactor-scale representations into a downstream microscopic mechanical response within the evaluated World Model.

This experiment therefore provides the strongest integration evidence in the suite: rather than demonstrating isolated competence at individual physical scales, it evaluates whether those representations can operate together as one connected computational system.

Result: PASS—91.3% full-chain success rate.

3.5.1 From Pairwise Integration to a Physical World Model

Tests 1 through 4 establish the individual computational and physical links required for cross-scale reasoning. Test 5 evaluates their composition.

Together, the five tests show a progression from shared computation to end-to-end physical integration:

shared domain-responsive computation
→ microscopic-to-macroscopic intervention
→ macroscopic-to-microscopic mechanical response
→ competing nonlinear multiphysics
→ complete molecule-to-reactor-to-cell integration

The resulting system is therefore not evaluated solely as a collection of independent scientific predictors. Under the declared evaluation contracts, it demonstrates the ability to propagate physically meaningful information across molecular, reactor, and biological-mechanical scales within one unified computational framework.

4 Differentiability and Cross-Scale Inverse Design

The preceding experiments evaluate the World Model primarily in the forward direction: given a physical intervention, can the model correctly propagate its consequences across scales?

Traditional computational physics can answer this question, but only through modular forward simulation. Physical states are serialized and transferred across rigid software boundaries—for example, from quantum chemistry into kinetic models and then into continuum CFD. These discrete hand-offs break the end-to-end computational graph, preventing gradients from propagating continuously across the full physical stack.

As a result, industrial optimization typically remains an iterative search problem:

$$\text{propose} \;\rightarrow\; \text{simulate} \;\rightarrow\; \text{evaluate} \;\rightarrow\; \text{repeat}.$$

Tests 1 through 5 establish that the World Model preserves meaningful physical relationships across heterogeneous scales, including bidirectional coupling, nonlinear competing-mechanism trade-offs, and complete molecule-to-reactor-to-cell integration. Because the architecture itself is end-to-end differentiable, these learned relationships can now be traversed in reverse.

This enables **Cross-Scale Inverse Design**.

Rather than asking only

$$x \;\rightarrow\; y$$

—what physical outcome follows from a proposed intervention—the model can optimize the inverse problem:

$$y^* \;\rightarrow\; x^*,$$

where y^* is a desired factory-scale objective and x^* may jointly include microscopic molecular variables and macroscopic process controls. In this evaluation, x_{micro} denotes controllable molecular and kinetic variables, including reaction enthalpy, activation barriers, and permitted ligand or formulation descriptors; x_{process} denotes reactor operating conditions; and x_{control} denotes time-dependent control inputs.

More generally, the model evaluates gradients of a macroscopic objective with respect to variables distributed across the physical hierarchy:

$$\nabla_{x_{\text{micro}}} J_{\text{macro}}, \qquad \nabla_{x_{\text{process}}} J_{\text{macro}}, \qquad \nabla_{x_{\text{control}}} J_{\text{macro}}.$$

The central question is therefore no longer whether the neural model is differentiable in software. It is whether these cross-scale gradients remain **physically actionable** when judged by independent classical physics.

4.1 Empirical Inverse-Design Performance & Gradient Actionability

To test this directly, we evaluated the World Model against a frozen modular reference workflow consisting of high-fidelity classical numerical solvers coupled to a SciPy SLSQP outer-loop optimizer.

The evaluation used a blind suite of **100 predefined industrial design targets**, comprising:

- **60 verified feasible targets**, requiring joint optimization of microscopic and macroscopic variables; and
- **40 intentionally infeasible targets**, constructed outside predefined physical bounds.

4.1.1 Actionable Cross-Scale Gradients

The most direct test of gradient usefulness is whether stepping in the neural optimization direction produces an actual improvement when the resulting candidate is evaluated by the classical verifier.

For each target, the model proposed a descent direction:

$$-\nabla_x J_{\text{neural}}.$$

That proposed intervention was then evaluated independently using the corresponding high-fidelity numerical workflow.

Across the blind evaluation set, **88% of proposed neural descent directions produced an improving objective response in the classical verifier**. An "improving objective response" is defined as a verifier-evaluated decrease in the frozen objective, $J(x_{\text{candidate}}) < J(x_{\text{baseline}})$, after applying the proposed intervention under the preregistered feasibility constraints.

This result demonstrates that the World Model's gradients are not merely mathematically valid derivatives of the neural surrogate. In the large majority of evaluated cases, they point toward interventions that also improve the independently computed physical objective.

4.1.2 Solver-Verified Feasible-Target Success

Across the **60 feasible targets**, which required joint optimization of molecular interventions and macroscopic process controls, the World Model generated solver-verified solutions for **55 targets**, achieving a:

$$\boxed{91.7\%}$$

feasible-target success rate.

4.1.3 Feasibility Discrimination

Inverse design must also recognize when a requested objective cannot be achieved within the available physical design space.

Among the **40 deliberately infeasible targets**, the model correctly rejected **38**, corresponding to a:

$$\boxed{95.0\%}$$

true-negative rejection rate.

These cases provide an explicit falsification test: rather than forcing a numerical answer for every requested objective, the model successfully identified most targets lying beyond the declared thermodynamic or transport constraints of the evaluated systems.

Together, these experiments test three increasingly demanding properties:

$$\text{useful gradient direction} \rightarrow \text{successful design} \rightarrow \text{recognition of physical impossibility.}$$

4.2 Solver-Call Reduction and Computational Acceleration

The value of cross-scale differentiability is not limited to faster neural inference. Its more important consequence is that it changes the structure of the optimization loop.

A conventional modular optimizer must repeatedly perturb candidate variables, execute expensive forward simulations, reconcile states across separate physical solvers, evaluate the resulting objective, and repeat until convergence.

The World Model instead provides a direct optimization direction through its differentiable cross-scale representation.

4.2.1 Neural Proposal Latency

Generating a neural cross-scale proposal required **38 milliseconds** of forward-pass latency on an NVIDIA H100.

For comparison, a single traversal of the legacy file-coupled modular orchestration pipeline required approximately **415 milliseconds** solely for process hand-offs, disk I/O, and state reconciliation, excluding the dominant cost of the underlying high-fidelity numerical solves.

4.2.2 End-to-End Optimization Acceleration

More significantly, the World Model bypassed a median of **124 classical solver calls per target** relative to the SLSQP-based modular optimization workflow.

As a result, median verified optimization time was reduced from:

$$> 500 \text{ hours}$$

to approximately:

$$\boxed{5.6 \text{ hours}},$$

including final high-fidelity classical verification.

The acceleration therefore does not primarily arise because a neural forward pass is faster than a classical solver. It arises because the model dramatically reduces the number of expensive forward simulations required to locate a viable design.

The resulting engineering workflow changes from:

$$\boxed{\text{guess} \rightarrow \text{simulate} \rightarrow \text{adjust} \rightarrow \text{repeat}}$$

to:

$$\boxed{\text{specify outcome} \rightarrow \text{generate intervention} \rightarrow \text{verify}}$$

These results demonstrate that the World Model is not merely a fast forward surrogate. By producing physically actionable cross-scale optimization directions, it reduces the number of expensive numerical design cycles required to reach solver-verified candidate solutions.

The remaining question is whether these generated interventions survive contact with the physical world. Section 5 therefore evaluates the model through prospective execution in held-out industrial environments.

5 Held-Out Industrial Validation

To establish that the foundation model's cross-scale causality and inverse-design capabilities generalize beyond *in silico* verification, the architecture was deployed in live, real-world manufacturing environments. The following cases represent prospective physical validations executed by industrial partners. In these evaluations, the model was tasked with generating optimized process parameters for previously untested regimes. These parameters were cryptographically frozen and timestamped, and subsequently executed in physical pilot plants.

5.1 Case 1: Specialty Chemical Batch-to-Continuous Scale-Up

Physical Context: The transition of chemical synthesis from batch manufacturing to continuous-flow processing is a highly desirable industrial objective, offering superior heat transfer, safety, and throughput. However, moving to continuous flow drastically alters the macroscopic transport phenomena and residence times, often leading to unpredicted side reactions.

The evaluation target was a highly exothermic liquid-phase substitution process. The industrial partner's historical batch configuration was constrained by a temperature-sensitive impurity pathway, resulting in an extensive 14-hour cycle time. The historical baseline for this process yielded an 82.4% isolated product and a 12.3% impurity profile.

World Model Intervention: Rather than relying on iterative, trial-and-error pilot testing to find a viable continuous-flow regime, the foundation model utilized its native gradient propagation to execute an inverse-design query. Tasked with maximizing target molecule yield while strictly bounding the thermodynamic impurity pathway, the model generated a novel continuous-flow operating window. The generated parameters explicitly defined the required steady-state macroscopic variables, including reactor temperature limits, fluid residence times, and active heat-removal conditions.

Physical Execution & Results: The model-generated operating window was cryptographically frozen and physically executed by the industrial partner in a continuous-flow pilot facility. The prospective physical run successfully stabilized the highly exothermic reaction, fundamentally altering the process economics.

- **Yield Maximization:** The steady-state isolated yield achieved under the model's recommended parameters was 96.7%, representing a 14.3 percentage-point absolute improvement over the historical batch baseline.
- **Impurity Suppression:** The temperature-sensitive side reaction was successfully suppressed, reducing the final impurity profile to 3.1% (a 9.2 percentage-

point absolute reduction), passing the partner's strict predefined $< 4.0\%$ acceptance criterion.

This prospective physical execution confirms that the foundation model can accurately compute the complex interplay between macroscopic continuous-flow transport and microscopic reaction kinetics, generating actionable industrial parameters that successfully transfer to physical reality.

5.2 Case 2: 500L Biomanufacturing Upstream & Downstream Scale-Up

Physical Context: Scaling up biomanufacturing processes—specifically for high-concentration monoclonal antibodies (mAbs)—presents a classic Macro-to-Micro multiphysics conflict. To ensure adequate oxygen mass transfer (k_La) in a large macroscopic reactor, operators must increase agitation. However, increased turbulent energy dissipation generates high hydrodynamic shear forces that can physically lyse (rupture) the microscopic mammalian cells.

The target objective was to scale a mAb fed-batch process from a 5-Liter benchtop bioreactor to a 500-Liter pilot facility, encompassing both upstream cell culture and downstream Tangential Flow Filtration (TFF) unit operations.

World Model Intervention: Demonstrating the capability to navigate non-linear competing mechanisms (as verified in Test 4), the foundation model generated a complete, cross-scale operating trajectory to thread the optimal boundary between oxygenation and cell viability. Upstream recommendations included dynamic impeller-speed profiles, sparger gas flow rates, and nutrient-feed strategies engineered to match the 5L baseline k_La while strictly bounding local shear stress. Downstream TFF parameters were inversely generated to optimize transmembrane pressure and crossflow flux without inducing product aggregation.

Physical Execution & Results: The generated parameters were cryptographically frozen and executed in a 500L physical pilot run, tracked by counter-signed batch and analytical records (summarized in Table 2).

- **Upstream Viability & Yield:** The process achieved a harvest product titer of 6.63 g/L (1.02× the 6.50 g/L target). Peak hydrodynamic shear stress was successfully bounded strictly below 0.15 N/m^2, effectively preventing shear-induced cell lysis and resulting in a harvest viability of 78.2% (satisfying the $> 75\%$ minimum specification for downstream clearance).
- **Downstream Filtration & Quality:** The generated TFF parameters achieved a 94.5% product recovery rate and a 14.5× concentration factor. Crucially, High Molecular Weight (HMW) aggregates were measured analytically at just 1.2%, comfortably below the strict $< 2.0\%$ safety threshold.

This execution provides physical proof that the model's internally computed gradients accurately optimize complex, multi-stage bioprocessing trajectories, successfully transferring an *in silico* design to a physical 500L facility.

5.3 Case 3: Formulation Stability & Live Continuous Optimization

Physical Context: While Cases 1 and 2 validated the model's ability to generate static setpoints for prospective pilot environments, industrial factories inherently experience continuous temporal deviations (e.g., raw-material feed fluctuations, asynchronous sensor delays). Furthermore, scaling novel biologic formulations requires linking microscopic electrostatic properties (isoelectric point and charge distribution) to macroscopic colloidal stability over months of physical shelf-life storage.

World Model Intervention & Results: The model was evaluated across two distinct commercial environments to test long-horizon stability and temporal anomaly resolution:

- **Continuous Root-Cause Optimization:** Deployed as a live inference engine on a commercial biomanufacturing line, the model ingested continuous, noisy industrial telemetry (pressure events, asynchronous at-line analytical data). By continuously correlating upstream mechanical and feed deviations directly to downstream titer impacts, the model provided dynamic state-corrections. Physical execution of these trajectory corrections improved overall upstream viability from 78% to 88% and increased target titer from 3.8 g/L to 4.5 g/L.
- **Accelerated Formulation Stability:** Operating on novel molecular formulations, the model's coupled quantum-to-mesoscopic adapters predicted isoelectric-point (pI) bounds and colloidal-stability trajectories. The 3-day result refers to accelerated computational screening against a conventional 12-week physical protocol; the corresponding 12-week stability outcome has not yet been prospectively confirmed. The model generated candidate formulation adjustments with predicted degradation decreasing from 8.5% to 2.5%, pending full-duration physical confirmation.

These deployments verify that the foundation model's cross-scale representations remain robust amid the noise, complexity, and extended temporal horizons of live industrial chemistry and biology.

6 Conclusion

The transition of advanced materials, chemicals, and biologics from laboratory discovery to industrial mass production has historically been constrained by the fragmentation of computational physics. Heterogeneous numerical solver stacks across atomistic, mesoscopic, and continuum regimes generally do not preserve a unified end-to-end differentiable path. Industrial scale-up therefore relies heavily on modular forward simulation and costly empirical trial-and-error, creating a "Valley of Death" in physical manufacturing.

In this report, we introduced a continuous foundation World Model for physical intelligence. By projecting domain-specific physical representations into a shared Sparse Mixture-of-Experts (MoE) architecture, the model

Table 2. 500L mAb Scale-Up (Physical Validation).

Scale-Up Parameter (5L → 500L)	**World Model Prediction / Target**	**Physical Execution Result**	**Status**
Harvest Titer	≥ 6.50 g/L	**6.63 g/L**	Verified
Cell Viability	$> 75.0\%$	**78.2%**	Verified
Peak Hydrodynamic Shear	< 0.15 N/m^2	**Bounded**	Verified
Product Recovery (TFF)	Maximized	**94.5%**	Verified
HMW Aggregates	$< 2.0\%$	**1.2%**	Verified

establishes a fully differentiable computational path across heterogeneous physical scales and representations.

We demonstrated bidirectional cross-scale capability: propagating microscopic quantum energetics upward to determine macroscopic reactor thermodynamics, and transferring macroscopic turbulent hydrodynamics downward to compute the structural failure of a microscopic biological cell. The model further represented nonlinear competing multiphysics regimes, capturing an operating landscape in which one physical response saturates while another continues to escalate. In the final integration test, these capabilities were composed into a continuous molecule-to-reactor-to-cell pathway with a **91.3% full-chain pass rate**.

Crucially, this differentiable representation enables **Cross-Scale Inverse Design**. In blind evaluation, **88% of proposed neural descent directions produced an improving objective response when independently evaluated by high-fidelity classical numerical solvers**. Across feasible inverse-design targets, the model achieved a **91.7% solver-verified success rate**, while prospective industrial executions demonstrated that generated interventions could successfully transfer from the computational design space into physical manufacturing environments—including a batch-to-continuous chemical scale-up that increased isolated yield from **82.4% to 96.7%**.

These results demonstrate that cross-scale gradients can remain physically actionable: the model does not merely predict the consequences of a proposed design, but can calculate useful directions through a heterogeneous physical design space toward a targeted outcome.

By unifying quantum thermodynamics, fluid dynamics, and biological mechanics within a shared differentiable computational framework, the foundation World Model begins to invert the conventional engineering paradigm. Rather than relying exclusively on humans to propose candidate processes for simulation, physical intelligence can increasingly generate the molecular interventions, operating conditions, and process controls required to achieve targeted physical and manufacturing outcomes at scale.

References

[1] D. S. Levine *et al.*, "The Open Molecules 2025 (OMol25) Dataset, Evaluations, and Models," arXiv:2505.08762 (2025).

[2] Z. Li, D. Z. Huang, B. Liu, and A. Anandkumar, "Fourier Neural Operator with Learned Deformations for PDEs on General Geometries," *Journal of Machine Learning Research* **24**, 1–26 (2023).

[3] M. Herde *et al.*, "Poseidon: Efficient Foundation Models for PDEs," *Advances in Neural Information Processing Systems* **37** (2024).

[4] J. Abramson *et al.*, "Accurate structure prediction of biomolecular interactions with AlphaFold 3," *Nature* **630**, 493–500 (2024).

[5] F. Neese, "Software update: the ORCA program system, version 5.0," *Wiley Interdisciplinary Reviews: Computational Molecular Science* **12**, e1606 (2022).

[6] H. G. Weller, G. Tabor, H. Jasak, and C. Fureby, "A tensorial approach to computational continuum mechanics using object-oriented techniques," *Computers in Physics* **12**, 620–631 (1998).

[7] N. Shazeer *et al.*, "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer," arXiv:1701.06538 (2017).

[8] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," arXiv:2312.00752 (2023).