HYPOTHESIZE, EVALUATE, REFINE: A SCIENTIFIC AGENT FOR OPEN-FORM PDE DISCOVERY WITH UNKNOWN SPATIAL COEFFICIENT FIELDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Understanding dynamics in heterogeneous media requires identifying both the governing differential operator and the spatial properties that determine its local action. Discovering these jointly is difficult: operator scope changes the law, while flexible coefficient fields can mask structural error. Existing open-form methods generally assume scalar or explicitly parameterized coefficients; nonparametric field estimators typically operate within prescribed operator libraries. We introduce Hypothesize, Evaluate, Refine for PDE Discovery (HER-PDE), coupling a scientific Agent with the Hypothesis Evaluation Interface (HEI). The Agent composes operators and field nodes into complete expression trees; HEI fits their time-invariant spatial fields and scores short-horizon prediction across two excitations within its supported evolution class. This separates structural reasoning from continuous estimation while preserving operator scope and shared-field identity. On seven two-dimensional heterogeneous PDE families with 5% state noise, including nonlinear cubic absorption, HER-PDE recovers 70/105 exact structures and achieves 94.3% Pass@3, compared with 39/105 exact recoveries for recursive typed evolution using the same HEI. Registered-structure fits attain median field correlation 0.891 across 65 fields; targeted controls test noise, solver, and morphology shifts. In a real spatial-drug case study, an Agent-proposed density-inhibitory effective model reduces continuous-prediction error by 15.7% relative to a refitted linear reference in post-discovery technical-repeat validation on measured microbial trajectories.

1 INTRODUCTION

Data-driven PDE discovery turns spatiotemporal observations into interpretable evolution laws for explanation and prediction (Rudy et al., 2017; Messenger & Bortz, 2021b; Du et al., 2024c). In aquifers, composite materials, and biological tissues, transport and reaction properties vary across space. The operator identifies the mechanism; coefficient fields describe where and how strongly it acts. Discovery must recover the governing structure and its unknown spatial coefficient fields.

A spatial field changes the law when moved across a differential operator:

$$c(\boldsymbol{x})\Delta u \quad \not\equiv \quad \nabla\cdot\big(c(\boldsymbol{x})\nabla u\big). \tag{1}$$

The second expression also contains $\nabla c\cdot\nabla u$. Field reuse and spatial dependencies likewise change the hypothesis. On a single trajectory, a flexible field can absorb operator error, coupling structural identification with estimation of the unknown spatial properties.

We build on the equation-discoverability framework of Lou et al. (2026), whose distinction between structural and coefficient complexity provides the conceptual basis for our positioning. Extending their literature survey, we curate and classify 87 equation-discovery publications (Figure 1; Section 2). Open-form methods construct expressions but generally retain scalar or explicitly parameterized coefficients (Chen et al., 2022; Du et al., 2024c; Xu et al., 2025a); nonparametric field estimators retain prescribed candidate terms (Luo et al., 2023; He et al., 2023; Tang et al., 2026). Our target joins these capabilities through open-form composition of operators and unknown *time-invariant spatial fields* $c_k(\boldsymbol{x})$. Here, open-form means constructing complete expressions without a predefined library of candidate terms; field shapes are estimated numerically without prescribing a compact formula. Time-varying and stochastic coefficient processes occupy separate regimes in the diagram.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

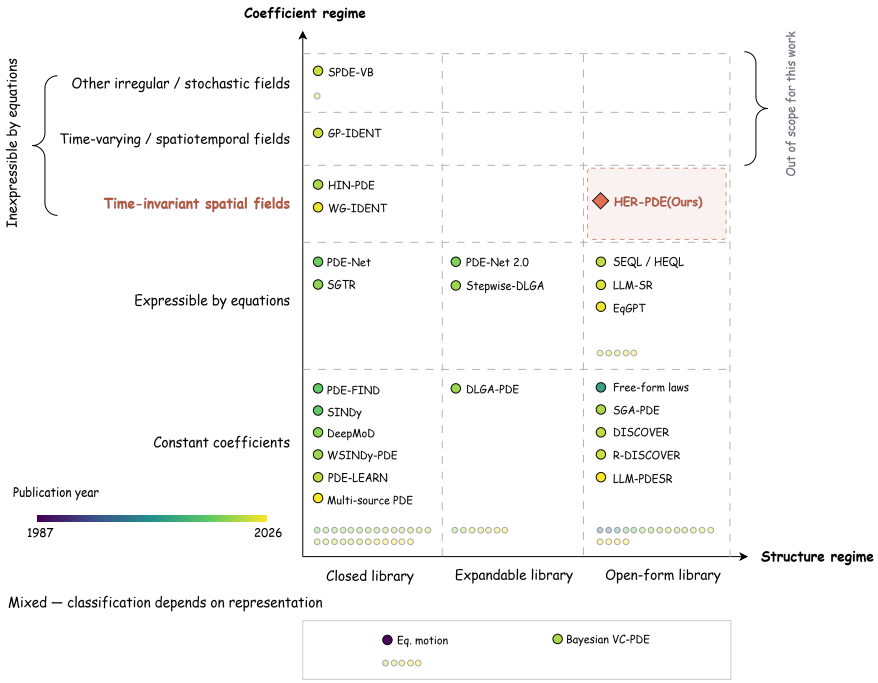


Figure 1: Equation-discoverability plane adapted and refined from Lou et al. (2026) through our extended 87-publication review. The horizontal axis describes structural freedom; the vertical axis describes coefficient expressiveness. Labels show representative methods from the audit (Appendix I); color denotes year, and within-cell placement is nonquantitative. HER-PDE occupies the previously empty intersection of open-form structure and unknown nonparametric spatial fields.

Large language models (LLMs) offer a useful search mechanism for this joint problem: they can combine experimental descriptions, executable data analysis, and numerical feedback when proposing symbolic explanations (Shojaee et al., 2025; Xia et al., 2026). This lets a scientific Agent reconsider operator placement, nonlinear composition, and field sharing using the accumulated evidence, while also refining promising candidates locally. Each proposal still requires a numerical test: an LLM's structural hypothesis does not determine the unknown coefficient fields.

We introduce *Hypothesize, Evaluate, Refine for PDE Discovery* (HER-PDE), coupling an LLM-based scientific Agent with a candidate-specific field evaluator. The Agent analyzes observations and constructs complete expression trees through global exploration and local refinement. The Hypothesis Evaluation Interface (HEI) tackles the continuous problem: it fits the fields declared by each tree jointly across two excitations and scores short-horizon prediction. The excitations expose different state and derivative combinations while the material fields remain invariant, turning cross-excitation consistency into structural evidence. The Agent supplies the complete law; HEI preserves its operator scope and field identity while solving the induced inverse problem.

We evaluate both synthetic and experimental systems. Seven two-dimensional PDE families isolate operator–field composition and nonlinear state dependence under known generating laws: HER-PDE recovers 70/105 exact structures against matched controls, with tests of noise, numerical solvers, and field morphology. A spatial-drug case study examines effective-model discovery from measured microbial dynamics. An Agent-proposed density-inhibitory model improves prediction over the refitted linear reference in a post-discovery comparison across technical repeats (Section 4.4).

Our contributions cover the discovery formulation, evaluator, Agent, and benchmark:

- We introduce the first framework identified in our curated 87-publication audit that jointly discovers open-form operator–field compositions and estimates unknown, time-invariant spatial fields, preserving operator scope, dependencies, and shared-field identity.

- We develop HEI to fit shared fields through differentiable short-horizon prediction and return predictive scores and field evidence for each complete hypothesis.
- We develop a two-phase scientific Agent that preserves one conversation while moving from raw-data understanding to complementary global hypothesis exploration and local refinement guided by candidate scores and fitted-field evidence.
- We introduce a seven-family, 35-environment evaluation protocol and report structural discovery over 105 runs, field-recovery and sealed-prediction metrics, controlled robustness tests, and a real-data effective-model case study of measured microbial growth and migration.

2 Related Work

Positioning by structure and coefficient space. Following Lou et al. (2026), Figure 1 separates structural freedom from coefficient complexity. Closed libraries select predefined terms, expandable libraries generate terms through restricted rules, and open-form methods recursively compose complete expressions. Vertically, constants and explicit coefficient formulas are distinguished from *Inexpressible by equations*: fields estimated without assuming a compact symbolic coefficient law. We refine this category into time-invariant spatial fields $c(\boldsymbol{x})$, time-varying or spatiotemporal fields $c(\boldsymbol{x}, t)$, and other irregular or stochastic coefficient models. HER-PDE targets the first of these with open-form operators. A randomly drawn spatial environment remains one fixed field during its trajectories; it is distinct from a coefficient process that changes over time or stochastic realizations. Numerical basis functions represent the field without prescribing its symbolic form.

Closed and expandable PDE libraries. Sparse identification selects active columns from a supplied feature matrix. SINDy introduced this formulation for dynamical systems and PDE-FIND extended it to PDEs (Brunton et al., 2016; Rudy et al., 2017). Weak forms and neural state surrogates improve derivative estimation (Messenger & Bortz, 2021b; Both et al., 2021). PDE-Net 2.0 learns differential filters and symbolic combinations (Long et al., 2019); DLGA-PDE expands terms by mutation and crossover (Xu et al., 2020). Construction rules set the allowed scope, depth, and coupling.

Open-form governing structures. Following free-form symbolic regression (Schmidt & Lipson, 2009), SGA-PDE searches forests of expression trees (Chen et al., 2022); DISCOVER and R-DISCOVER generate trees by reinforcement learning (Du et al., 2024c;a); and EqGPT learns priors from handbook equations (Xu et al., 2025a). These methods compose operators beyond a fixed matrix but do not attach unknown nonparametric fields at candidate-specific tree positions. HER-PDE makes those fields explicit expression nodes with shared identities across the tree.

Variable-coefficient discovery. SGTR, GP-IDENT, and Bayesian group-sparse methods estimate coefficient profiles within supplied dictionaries (Rudy et al., 2019; He et al., 2023; Chen et al., 2026). Stepwise-DLGA and SEQL/HEQL seek compact coefficient formulas (Xu et al., 2021; Zhang et al., 2024). HIN-PDE combines a fixed library with Gaussian-kernel fields (Luo et al., 2023); WG-IDENT uses weak forms, spline fields, and group selection (Tang et al., 2026). In HER-PDE, each tree instead determines which fields exist, where they occur, and which occurrences share one function. Its kernel basis estimates candidate-specific spatial field values, not a predefined term support.

Agents for scientific and equation discovery. LLM4ED uses numerical feedback (Du et al., 2024b), LLM-SR optimizes generated program skeletons (Shojaee et al., 2025), and SR-Scientist combines executable analysis with long-horizon symbolic search (Xia et al., 2026). Related agents use simulators and optimizers (Ma et al., 2024a). HER-PDE couples this reasoning loop to the nonparametric inverse problem induced by each complete PDE tree.

Appendix I gives the classification rubric and all 87 publications underlying Figure 1.

3 Method

HER-PDE separates structural search from field estimation. The LLM-based Scientific Agent proposes complete laws; HEI fits each law's unknown fields and returns predictive evidence. Figure 2 shows the workspace, tools, and two-phase hypothesis-search procedure.

3.1 Open-form PDEs with unknown spatial fields

Input and discovery target. We observe a scalar state $u(\boldsymbol{x}, t)$ over a spatial domain $\Omega \subset \mathbb{R}^d$. Two noisy measurement sequences $\widetilde{u}^{(r)}(\boldsymbol{x}_i, t_n)$, $r \in \{A, B\}$, come from the same system and share its

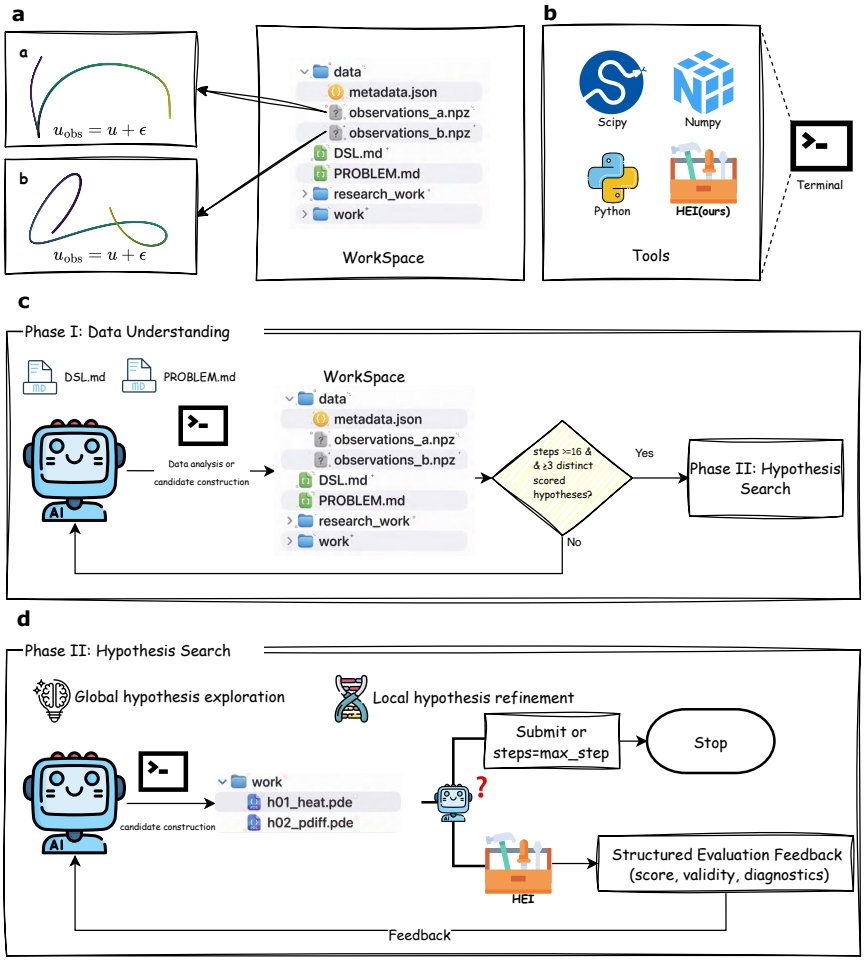


Figure 2: HER-PDE couples scientific hypothesis generation with candidate-specific field evaluation. (a) The workspace contains two observed trajectories, experimental context (`PROBLEM.md`), and the candidate grammar (`DSL.md`). (b) Terminal tools support analysis and candidate construction; HEI fits declared fields and returns predictive scores and field evidence. (c) Phase I analyzes data and scores initial hypotheses; transition requires at least 16 turns and three distinct scored candidates. (d) The same conversation continues with raw arrays hidden. Global exploration constructs alternative explanations, while local refinement modifies scored candidates. Repeated HEI feedback guides selection of one final PDE.

time-invariant coefficient fields. Our goal is to recover a complete symbolic evolution law together with the spatial functions that parameterize its local action throughout the observed domain:

$$\partial_t u = \mathcal{F}_\xi[u; c_0, \ldots, c_{K-1}](\boldsymbol{x}, t), \qquad c_k : \Omega \to \mathbb{R}, \tag{2}$$

The expression tree ξ determines the operators, their composition, the number K of unknown fields, and where each field is used. The functions c_k may vary with position but remain fixed across time and both sequences. Different excitations test the same fields against distinct state histories.

Complete expression-tree hypotheses. A hypothesis $h = (\xi, \mathcal{C})$ pairs the tree with declarations $\mathcal{C} = \{(C_k, S_k)\}_{k=0}^{K-1}$. Here C_k names a field and $S_k \subseteq \{x_1, \ldots, x_d\}$ lists its spatial arguments, abbreviated $\boldsymbol{x}_{S_k}$. The grammar recursively combines the state, coordinates, arithmetic, pointwise functions, derivatives, and these field nodes. It specifies how to construct expressions rather than listing complete candidate terms. Reusing a name ties all occurrences to one unknown function, including occurrences inside derivatives. Different names declare independent fields. Discovery returns this complete tree and its fitted fields, preserving operator scope and shared-field identity.

3.2 HYPOTHESIS EVALUATION INTERFACE

HEI evaluates autonomous scalar evolution laws with nonlinear state dependence and affine dependence on unknown fields and their derivatives after differentiation (Appendix Eq. (8)). It fits only the declared fields, returning validity, predictive scores, and fitted-field evidence. Appendix A specifies the supported spatial derivatives, field signatures, and boundary discretization.

Preparing the observed states. HEI starts from the supplied measurements. Coordinates reconstruct the grid; fixed preprocessing $\mathcal{S}$ then produces $u_{\mathrm{p}}^{(r)} = \mathcal{S}(\widetilde{u}^{(r)})$ for numerical evaluation. In the noisy synthetic benchmark, spatial Savitzky–Golay smoothing reduces measurement noise in each snapshot (Savitzky & Golay, 1964); the drug measurements use the original values. This preparation is identical for all candidates on a dataset. The resulting states supply both the starting observations and comparison targets for prediction, without estimating a temporal derivative.

Nonparametric fields. For every declared field $c_k(\boldsymbol{x}_{S_k})$, coordinates are normalized to the unit box and a tensor grid of anchors $\{\boldsymbol{a}_{kj}\}_{j=1}^{M_k}$ is formed from equally spaced coordinate quantiles, including the observed coordinate extrema. HEI uses the row-normalized Gaussian basis

$$\kappa_\sigma(\boldsymbol{z}, \boldsymbol{a}_{kj}) = \frac{\exp\!\left(-\|\boldsymbol{z} - \boldsymbol{a}_{kj}\|_2^2/(2\sigma^2)\right)}{\sum_{\ell=1}^{M_k} \exp(-\|\boldsymbol{z} - \boldsymbol{a}_{k\ell}\|_2^2/(2\sigma^2))}, \tag{3}$$

$$c_k(\boldsymbol{z}; \boldsymbol{\beta}_k) = \sum_{j=1}^{M_k} \beta_{kj} \kappa_\sigma(\boldsymbol{z}, \boldsymbol{a}_{kj}). \tag{4}$$

One weight vector $\boldsymbol{\beta}_k$ determines every occurrence of field c_k. In particular, a field and its derivative are tied: if B_k evaluates the basis on the grid and D_q is the physical-grid difference operator, their grid values are $\boldsymbol{c}_k = B_k\boldsymbol{\beta}_k$ and $(\partial_{x_q} c_k)_{\mathrm{grid}} = \mathsf{D}_q B_k \boldsymbol{\beta}_k$. There is no separate weight vector for a derivative of the same field. Recognized conservative products use face-flux discretization with that same field (Appendix A.2). We adapt HIN-PDE's Gaussian-kernel field representation (Luo et al., 2023) and fit its weights within the inverse problem defined by each submitted tree.

Candidate-consistent state prediction. The compiler applies differentiation rules to the submitted tree while preserving its field identities. For each trial field estimate, it integrates the candidate forward from an observed state and compares the prediction with a later observation. At each integration stage, the candidate's discrete operators act on the evolving state using the same fitted fields. Let Φ_h^H denote classical fourth-order Runge–Kutta integration of the complete candidate across H observed time intervals. For transition starts $\mathcal{I}$ in trajectory r, let Ω° be the scored interior grid and let $\overline{u}_{\mathrm{p},\mathcal{I},H}^{(r)}$ be the scalar mean of $u_{\mathrm{p},n+H}^{(r)}(\boldsymbol{x}_i)$ over all $(n, \boldsymbol{x}_i) \in \mathcal{I} \times \Omega^\circ$. HEI minimizes the normalized prediction error over the selected transition starts and spatial points:

$$\mathcal{L}_r(\mathcal{I}; h, \boldsymbol{\beta}, \sigma) = \frac{\mathrm{mean}_{n\in\mathcal{I},\, \boldsymbol{x}_i\in\Omega^\circ}\left[u_{\mathrm{p},n+H}^{(r)}(\boldsymbol{x}_i) - \Phi_h^H(u_{\mathrm{p},n}^{(r)}; \boldsymbol{\beta}, \sigma)(\boldsymbol{x}_i)\right]^2}{\max\left\{\mathrm{mean}_{n\in\mathcal{I},\, \boldsymbol{x}_i\in\Omega^\circ}\left[u_{\mathrm{p},n+H}^{(r)}(\boldsymbol{x}_i) - \overline{u}_{\mathrm{p},\mathcal{I},H}^{(r)}\right]^2,\ 10^{-12}\right\}}. \tag{5}$$

The score measures state-prediction error normalized by observed target variance. The integration horizon, scored grid region, boundary discretization, and numerical integration settings are fixed for all candidates within each observation protocol (Appendices A.2 and G).

The time intervals are partitioned chronologically into fit, development, reward, and sealed-final blocks with proportions $40/20/20/20$. For every kernel width, ridge regression on a time-integrated linear surrogate induced by Eq. (8) initializes the field coefficients. Gradient-based optimization then minimizes the complete RK4 prediction loss over A and B using one shared coefficient vector. The width is chosen by mean A/B development loss; the selected field is refitted on fit plus development. The reward-block error supplies the score returned to the Agent:

$$s(h) = \tfrac{1}{2}\left[\mathcal{L}_A(\mathcal{I}_{\mathrm{reward}}; h, \widehat{\boldsymbol{\beta}}, \widehat{\sigma}) + \mathcal{L}_B(\mathcal{I}_{\mathrm{reward}}; h, \widehat{\boldsymbol{\beta}}, \widehat{\sigma})\right]. \tag{6}$$

Sharing one field across both sequences prevents trajectory-specific compensating laws. The sealed block is evaluated only after the Agent submits its final candidate.

Agent-visible interface. The Agent invokes six terminal operations: `validate`, `score`, `score-batch`, `history`, `evidence`, and `submit`. They expose candidate validity, predictive ranking, fitted-field evidence, and finalization without revealing sealed data or modifying a submitted tree. Appendix Table 3 defines the complete return contract.

Table 1: Registered PDE families and the structural capability isolated by each family. All coefficient symbols denote unknown fields of (x, y) shared by both observed trajectories.

Family	Registered law	Structural capability
Variable diffusion	$u_t = D(u_{xx} + u_{yy})$	Shared field across directions
Conservative diffusion	$u_t = \partial_x(Du_x) + \partial_y(Du_y)$	Operator scope
Reaction–diffusion	$u_t = \nabla\cdot(D\nabla u) - ku$	Operator-specific fields
Advection–diffusion	$u_t = D\Delta u - v_x u_x - v_y u_y$	Three independent fields
Orthotropic diffusion	$u_t = \partial_x(D_x u_x) + \partial_y(D_y u_y)$	Directional flux fields
Pure advection	$u_t = -v_x u_x - v_y u_y$	Diffusion-free first order law
Cubic absorption	$u_t = D\Delta u - ku^3$	Nonlinear state, two fields

3.3 Two-phase scientific Agent

One conversation connects data analysis with hypothesis tests. Its workspace contains `PROBLEM.md` (observation setting and data), `DSL.md` (domain-specific candidate language and HEI commands), and the measurement arrays with coordinates. The reference equation is not supplied. Through a terminal, it runs Python, NumPy, and SciPy, writes analysis code, and saves complete equations as `work/*.pde`. Calling `pde score work/<name>.pde` returns validity, predictive errors, and an archive identifier; `evidence` exposes fitted fields and `history` compares candidates.

Phase I: data understanding. Phase I establishes what variation the data support (Figure 2c). During the first 16 model turns, it may inspect arrays, reconstruct the grid, study sampling and noise, and write numerical diagnostics. Candidate construction and scoring are available throughout. A checkpoint requires at least three distinct scored hypotheses, directing subsequent actions to new scored proposals until that minimum is reached. This converts exploratory analysis into an initial set of competing explanations and numerical evidence for subsequent structural search.

Phase II: hypothesis search. Phase II focuses the remaining budget on testing explanations (Figure 2d). Raw arrays are removed from the workspace, while the same conversation retains analysis results, code, candidate files, the score archive, and fitted-field evidence. This redirects tool use from raw-data analysis to hypothesis comparison without losing scientific context. Global hypothesis exploration and local hypothesis refinement guide the search:

Global hypothesis exploration returns to accumulated evidence and constructs an alternative explanation without editing a scored parent. It reconsiders operator composition, derivative scope, nonlinearity, and field sharing, challenging the assumptions behind the current leading candidate.

Local hypothesis refinement selects a scored parent, identifies an evidence-based weakness, and makes one controlled structural change. It then scores the child and compares prediction and field evidence with the parent. Phase II starts globally and requires both operations before selection. The Agent chooses their order to explore alternatives and refine promising candidates.

Selection and stopping. The budget is 36 model turns and at most 96 unique scores. If no submission has occurred by the final model turn, two bounded selection turns permit only archive inspection and submission. The Agent selects one scored candidate using prediction, field evidence, structural coherence, and unsupported complexity; HEI then evaluates the sealed block. Appendix E reproduces the exact system, checkpoint, search, and selection prompts used in the frozen runs.

4 Experiments

4.1 Frozen protocol

PDE families. Table 1 isolates seven capabilities in two-dimensional scalar PDEs, including nonlinear state dependence. Coefficients are unknown, time-invariant functions of (x, y) that enter linearly; their forms and values remain hidden from the Agent.

Environments and observations. Each family has five independently seeded Matérn-$3/2$ environments: smooth, medium, rough, anisotropic, and two-scale. Positive diffusion and reaction fields are log-Gaussian transforms; signed velocity fields retain both signs. Each supplies two excitations of the same material on a 49×49 grid at 41 times, with 5% relative Gaussian state noise. The Agent

Table 2: Structural discovery over 35 environments and 105 runs per method. Exact includes Pass@1; invalid finishes count as failures. † marks the 96-evaluation cap reached in every run; N/A denotes no HEI. "Sep." fits A/B separately. Family results: Appendix Table 10.

Method	Field fit	HEI evals per run	Exact runs (Pass@1 %)	Pass@2 (%)	Pass@3 (%)
OpenHands	None	N/A	4/105 (3.8)	6.7	8.6
Typed evolution	Shared	96†	39/105 (37.1)	45.7	51.4
Flat Agent	Shared	13.0	27/105 (25.7)	42.9	54.3
HER-PDE (Sep.)	Separate	31.2	62/105 (59.0)	73.3	80.0
HER-PDE	Shared	29.6	**70/105 (66.7)**	**88.6**	**94.3**

sees coordinates, times, and noisy states through index 32; the final eight states and all simulator-side equations, fields, clean states, derivatives, and conditions remain hidden.

The original suite and cubic extension freeze generation before their formal runs, use seeds disjoint from development, and share the same Agent and HEI settings. All registered environments are retained. Appendices B and F specify data generation and cohort accounting.

Agent and compute. We use the DeepSeek-V4-Flash-0731 API model[1] (DeepSeek-AI et al., 2026) with high-effort thinking. Three API rollouts per environment give 105 runs. The frozen two-dimensional evaluator uses full spatial field signatures, 12×12 Gaussian anchors, and autonomous laws with diagonal spatial second derivatives. Spatial smoothing uses a window of seven and polynomial degree three. Four-interval predictions use eight RK4 substeps per interval and score the grid interior. Appendix A.2 specifies boundary rows and optimization settings.

Structural metric. A run succeeds when its submission is algebraically equivalent to the registered law up to field renaming, preserving dependencies, scope, shared-field identity, and term support.

Equation discovery produces hypotheses for scientific validation: a useful search can place the governing law in a small shortlist even when a single run misses it. Pass@k measures this opportunity against the cost of checking up to k final laws from independent runs. Pass@1 is the order-invariant single-run success estimate; with three runs per environment, it equals the pooled exact-run rate. Pass@2 and Pass@3 measure whether a two- or three-run shortlist contains the registered law. Appendix Eq. (18) defines the estimator (Chen et al., 2021a) and environment-level bootstrap.

Field and prediction metrics. Pearson correlation and relative L_2 compare fitted and generating fields at the HEI anchors. Sealed prediction uses Eq. (5). Appendix B.8 defines alignment and cohorts.

4.2 Structural discovery and controlled comparisons

Same-backbone OpenHands (Wang et al., 2025) is a system-level comparison with data-analysis tools but no HEI. Typed evolution (Xu et al., 2020; Chen et al., 2022; Fortin et al., 2012) controls non-LLM search with a bounded recursive grammar and the same HEI. Flat Agent is an unstaged LLM loop with numerical feedback, matching the backbone, candidate language, HEI, and turn/score ceilings. A separate-field variant retains the full Agent but fits A and B independently. Appendix C specifies the matched interventions, evaluation protocols, and budget ceilings.

With the same LLM and HEI, HER-PDE recovers 70/105 structures versus 27/105 for Flat Agent (Table 2). The whole-controller comparison holds budget ceilings fixed and includes allocation of analysis and candidate testing: mean HEI use is 29.6 versus 13.0 evaluations per run.

Typed evolution reaches 39/105 while exhausting all 96 evaluations per run; HER-PDE achieves higher recovery with fewer evaluations and covers 33/35 environments within three attempts. Appendix C gives family-level comparisons, submission completion, and resource accounting.

Fixed-library regression. A GP-IDENT-style control (He et al., 2023) receives all target expansions in a 19-term library, assigning an independent 144-parameter field to each term. One fit per

[1]The 0731 release is documented in the 31 July 2026 entry of the official API change log: https://api-docs.deepseek.com/updates/#date-2026-07-31.

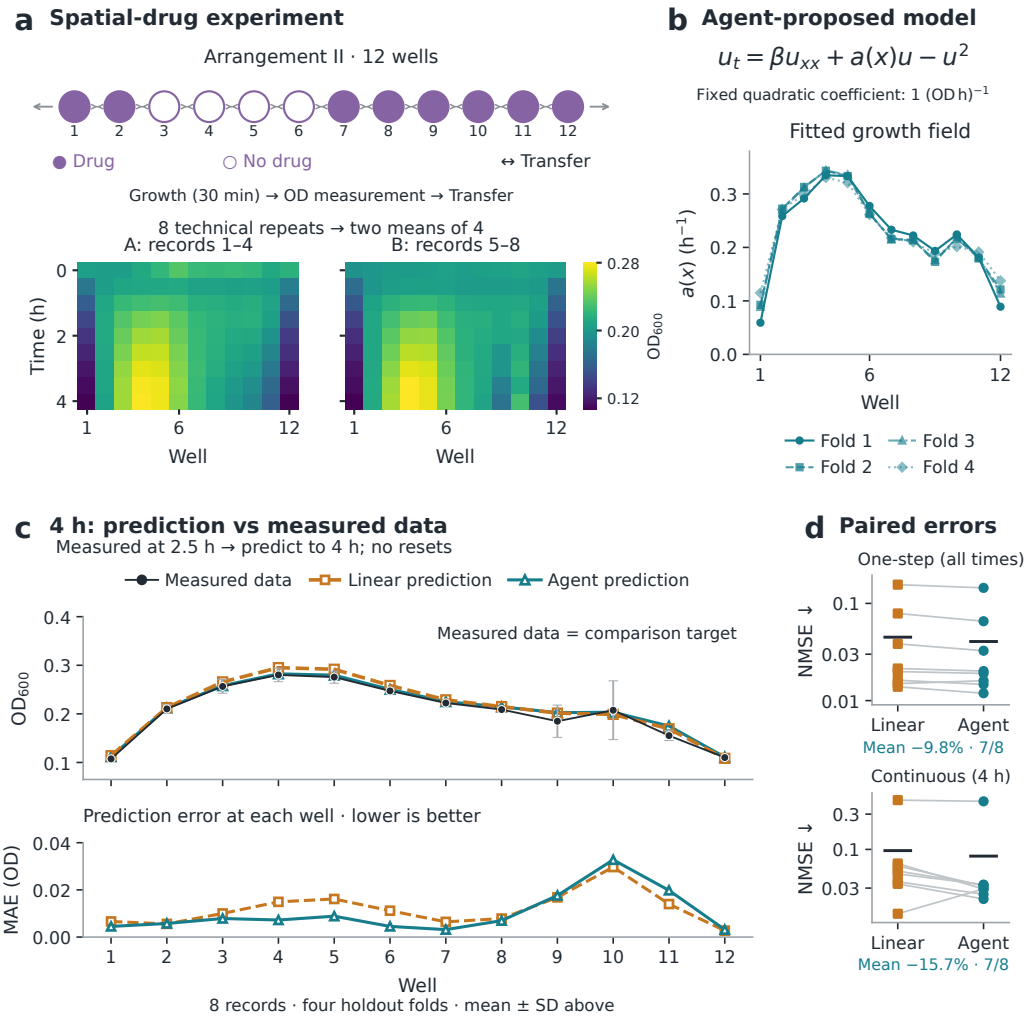


Figure 3: Effective-model case study on spatial-drug arrangement II. (a) Eight technical repeats in twelve wells form two disjoint four-record discovery means. (b) The Agent-proposed density-inhibitory model and four post-discovery refits of $a(x)$, each excluding two records from parameter fitting. (c) The proposed model and refitted linear reference predict from each excluded record's measured state at 2.5 h to 4 h without resets. Measured OD is shown as mean $\pm$ sample SD over eight records; the lower plot gives per-well MAE. (d) Lines pair record-level NMSE for half-hour one-step and continuous terminal predictions; black strokes mark means. Both models use matched refits and absorbing boundaries. The proposed model reduces mean continuous terminal NMSE by 15.7%.

environment recovers 7/35 exact expanded supports, with micro-TPR and micro-PPV both 81.7%. Compositional and shared-field-identity recovery are N/A for this hypothesis class (Appendix C.1).

Shared fitting improves exact recovery from 62/105 to 70/105 and Pass@3 from 80.0% to 94.3%, with gains concentrated in reaction and transport; separate fitting selects more diffusion structures. With conservative diffusion fixed, sharing instead improves median field correlation from 0.898 to 0.955 and relative L_2 error from 0.143 to 0.097. Appendix C.2 gives the complete paired analysis.

Field recovery and sealed prediction. The 65 oracle-structure fields attain median correlation 0.891 and relative L_2 error 0.232. The diffusion and absorption fields in cubic absorption attain $0.852/0.212$ and $0.403/0.816$ (correlation/error), respectively. Across 104 submissions, median sealed error is 7.11×10^{-4} for exact structures and 2.01×10^{-3} for non-hits. Appendix B.8 breaks down these metrics by PDE family, physical operator, and fitting cohort.

4.3 Robustness to noise, numerical, and coefficient shifts

With the Agent and HEI fixed, three representative families yield 30/45, 30/45, and 32/45 exact structures at 1%, 5%, and 10% noise. On two diffusion families, an independent fine-grid generator yields 9/10 exact structures, while discontinuous or channelized fields yield 32/36. With the target structure fixed, HEI recovers 12 morphology-shift fields with median Pearson correlation 0.849 and relative L_2 error 0.172. Appendices B.6 and B.7 give protocols, tables, and uncertainty intervals.

4.4 Effective-model discovery in a spatial drug landscape

Experimental system and discovery inputs. Hu et al. (2026) grew *Enterococcus faecalis* in a row of twelve wells to study how spatial antibiotic exposure affects population spread. Each half-hour growth period was followed by robotic transfer between neighboring wells; removal at the two ends represented population loss to an absorbing exterior. A plate reader measured optical density at 600 nm (OD_{600}), a proxy for bacterial density. We use arrangement II: wells 3–6 are drug-free and the other eight contain linezolid. Eight technical repeats provide nine observations over four hours. Disjoint means of records 1–4 and 5–8 form inputs A/B (Figure 3a). The Agent receives the measurement context and arrays without the drug map or reference equation; HEI incorporates the known exterior loss. Three runs produce the submissions in Appendix G; the first is

$$u_t = \beta u_{xx} + a(x)u - u^2, \tag{7}$$

where u is measured optical density, β is a fitted constant, and $a(x)$ is an unknown spatial field.

Interpretable density inhibition. The term βu_{xx} represents migration, while $a(x)u - u^2$ combines a spatial growth landscape with density-dependent inhibition. Its per-capita growth decreases with OD, as in heterogeneous logistic population models (Nelson & Shnerb, 1998). Hu et al.'s reference model is the low-density linearization of Fisher–KPP, explicitly neglecting intraspecies competition (Hu et al., 2026, Supplementary Text, Section 1). The submitted term therefore has an ecological interpretation as a finite-density correction.

We compare this submission with that linear structure using matched, convergence-checked post-discovery refits on two three-record means per fold. Across the four refits, the growth intercept $a(x)$ peaks inside the drug-free block and falls toward the ends, matching the experimental arrangement (Figure 3b).

Both models reproduce the broad observed OD profile, so their mean predictions nearly overlap (Figure 3c). The quadratic model lowers per-well MAE around wells 3–6; both retain a larger residual near well 10.

Figure 3d shows improvements on seven of eight records for both protocols. Mean NMSE falls from 0.0963 to 0.0812 (15.7%) for continuous terminal prediction and from 0.0449 to 0.0405 (9.8%) for half-hour one-step forecasts reset to observations. This is post-discovery technical-repeat validation: all records contributed to the earlier discovery means. Appendix G gives the boundary, objective, all submissions, and sensitivity to fitting a free quadratic coefficient.

5 Conclusion

HER-PDE unifies scientific hypothesis generation and candidate-faithful field estimation. Across seven heterogeneous PDE families, it recovers 70/105 exact structures and covers 33/35 environments within three attempts, using fewer evaluations than typed evolution. A real spatial-drug case further demonstrates effective-model discovery through improved post-discovery prediction.

AI use statement

Generative AI assisted retrieval and discovery of related work, research ideation and experimental design, method implementation, result analysis and interpretation, figure preparation, and drafting and revising manuscript sections. It also assisted in developing the numerical PDE simulation code used to generate the synthetic datasets. The authors reviewed the AI-assisted work against the code, experiment configurations, result records, and cited sources, and take responsibility for the scientific accuracy and integrity of the manuscript.

Reproducibility statement

Appendices A–E and G specify the method, benchmarks, prompts, resampling, and real-data validation. Versioned configurations, seeds, and immutable result ledgers accompany the code.

References

E Paulo Alves and Frederico Fiuza. Data-driven discovery of reduced plasma physics models from fully kinetic simulations. *Physical Review Research*, 4(3):033192, 2022.

Hassan Arbabi, Judith Bunder, Giovanni Samaey, A. Roberts, and Yannis Kevrekidis. Linking machine learning with multiscale numerics: Data-driven discovery of homogenized equations. *JOM*, 72:4444–4457, 10 2020.

Francesco Ballerin and Erlend Grong. Discovering pdes equivariant under rigid motions, 2026. URL https://arxiv.org/abs/2608.08913.

Josh Bongard and Hod Lipson. Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 104(24):9943–9948, 2007.

Gert-Jan Both, Subham Choudhury, Pierre Sens, and Remy Kusters. Deepmod: Deep learning for model discovery in noisy data. *Journal of Computational Physics*, 428:109985, 2021. ISSN 0021-9991.

Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

Hongqing Cao, Lishan Kang, Yuping Chen, and Jingxian Yu. Evolutionary modeling of systems of ordinary differential equations with genetic programming. *Genetic Programming and Evolvable Machines*, 1(4):309–337, 2000.

Aoxue Chen, Yifan Du, Liyao Mars Gao, and Guang Lin. Bayesian data-driven discovery of partial differential equations with variable coefficients. *Computers & Mathematics with Applications*, 220:371–385, 2026. ISSN 0898-1221.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgen Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N. Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. Evaluating large language models trained on code, 2021a. URL https://arxiv.org/abs/2107.03374.

Yuntian Chen, Yingtao Luo, Qiang Liu, Hao Xu, and Dongxiao Zhang. Symbolic genetic algorithm for discovering open-form partial differential equations (sga-pde). *Phys. Rev. Res.*, 4:023174, Jun 2022.

Zhao Chen, Yang Liu, and Hao Sun. Physics-informed learning of governing equations from scarce data. *Nature communications*, 12(1):6136, 2021b.

Benjamin G Cohen, Burcu Beykal, and George M Bollas. Physics-informed genetic programming for discovery of partial differential equations from scarce and noisy data. *Journal of Computational Physics*, 514:113261, 2024.

Kevin Course and Prasanth Nair. State estimation of a physical system with unknown governing equations. *Nature*, 622:261–267, 10 2023.

Miles Cranmer, Alvaro Sanchez Gonzalez, Peter Battaglia, Rui Xu, Kyle Cranmer, David Spergel, and Shirley Ho. Discovering symbolic models from deep learning with inductive biases. *Advances in neural information processing systems*, 33:17429–17442, 2020.

J Crutchfield and B McNamara. Equations of motion from a data series. *Complex systems*, 1:417–452, 1987.

Stéphane d'Ascoli, Sören Becker, Philippe Schwaller, Alexander Mathis, and Niki Kilbertus. Odeformer: Symbolic regression of dynamical systems with transformers. In *International Conference on Learning Representations*, volume 2024, pp. 21943–21976, 2024.

Miguel Ángel de Carvalho Servia, King Kuok Hii, Klaus Hellgardt, Dongda Zhang, and Ehecatl Antonio del Rio Chanona. Simplest Mechanism Builder Algorithm (SiMBA): An Automated Microkinetic Model Discovery Tool. *Chemical Science*, 17:7533–7550, 2026.

DeepSeek-AI, Anyi Xu, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, Chenchen Ling, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chengyu Hou, Chenhao Xu, Chenze Shao, Chong Ruan, Conner Sun, Damai Dai, Daya Guo, Dejian Yang, Deli Chen, Donghao Li, Dongjie Ji, Erhang Li, Fang Wei, Fangyun Lin, Fangzhou Yuan, Feiyu Xia, Fucong Dai, Guangbo Hao, Guanting Chen, Guoai Cao, Guolai Meng, Guowei Li, Han Yu, Han Zhang, Hanwei Xu, Hao Li, Haofen Liang, Haoling Zhang, Haoming Luo, Haoran Wei, Haotian Yuan, Haowei Zhang, Haowen Luo, Haoyu Chen, Haozhe Ji, Hengqing Zhang, Honghui Ding, Hongxuan Tang, Huanqi Cao, Huazuo Gao, Hui Qu, Hui Zeng, J Yang, JQ Zhu, Jia Luo, Jia Song, Jia Yu, Jialiang Huang, Jialu Cai, Jian Liang, Jiangting Zhou, Jiasheng Ye, Jiashi Li, Jiaxin Xu, Jiewen Hu, Jieyu Yang, Jin Chen, Jin Yan, Jingchang Chen, Jingli Zhou, Jingting Xiang, Jingyang Yuan, Jingyuan Cheng, Jingzi Zhou, Jinhua Zhu, Jiping Yu, Joseph Sun, Jun Ran, Junguang Jiang, Junjie Qiu, Junlong Li, Junmin Zheng, Junxiao Song, Kai Dong, Kaige Gao, Kang Guan, Kexing Zhou, Kezhao Huang, Kuai Yu, Lean Wang, Lecong Zhang, Lei Wang, Leyi Xia, Li Zhang, Liang Zhao, Lihua Guo, Lingxiao Luo, Linwang Ma, Linyan Zhu, Litong Wang, Liyu Cai, Liyue Zhang, Longhao Chen, MS Di, MY Xu, Max Mei, Miaojun Wang, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Mingming Li, Mingxu Zhou, Minmin Han, Ning Wang, Panpan Huang, Panpan Wang, Peixin Cong, Peiyi Wang, Peng Zhang, Qiancheng Wang, Qihao Zhu, Qingyang Li, Qinyu Chen, Qiushi Du, Qiwei Jiang, Rui Tian, Ruifan Xu, Ruijie Lu, Ruiling Xu, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, Runqian Chen, Runqiu Yin, Runxin Xu, Ruomeng Shen, Ruoyu Zhang, Ruyi Chen, SH Liu, Shanghao Lu, Shangmian Sun, Shangyan Zhou, Shanhuang Chen, Shaofei Cai, Shaoheng Nie, Shaoqing Wu, Shaoyuan Chen, Shengding Hu, Shengyu Liu, Shiqiang Hu, Shirong Ma, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, Shuying Yu, Songyang Zhou, Tao Ni, Tao Yun, Tian Jin, Tian Pei, Tian Ye, Tianle Lin, Tianran Ji, Tianyi Cui, Tianyuan Yue, Tingting Yu, Tun Wang, W Zhang, WL Xiao, Wangding Zeng, Wei An, Weilin Zhao, Wen Liu, Wenfeng Liang, Wenjie Pang, Wenjing Luo, Wenjing Yao, Wenjun Gao, Wenkai Yang, Wenlve Huang, Wenqing Hou, Wentao Zhang, Wenting Ma, Xi Gao, Xiang He, Xiangwen Wang, Xianzu Wang, Xiao Bi, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaokang Zhang, Xiaotao Nie, Xiaowen Sun, Xiaoxiang Wang, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xingchen Liu, Xingkai Yu, Xingyou Li, Xinyu Yang, Xinyu Zhang, Xu Chen, Xuanyu Wang, Xuecheng Su, Xueyin Chen, Xuheng Lin, Xuwei Fu, YC Yan, YQ Wang, YW Ma, Yanfeng Luo, Yang Zhang, Yanhong Xu, Yanru Ma, Yanwen Huang, Yao Li, Yao Li, Yao Xu, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Qian, Yi Shao, Yi Yu, Yichao Zhang, Yifan Ding, Yifan Shi, Yijia Wu, Yiliang Xiong, Yiling Ma, Ying He, Ying Tang, Ying Zhou, Yingjia Luo, Yinmin Zhong, Yishi Piao, Yisong Wang, Yixiang Zhang, Yixiao Chen, Yixuan Tan, Yixuan Wei, Yiyang Ma, Yiyuan Liu, Yonglun Yang, Yongqiang Guo, Yongtong Wu, Yu Wu, YuKun Li, Yuan Cheng, Yuan Ou, Yuanfan Xu, Yuanhao Li, Yuduan Wang, Yuehan Yang, Yuer Xu, Yuhan Wu, Yuhao Meng, Yuheng Zou, Yukun Zha, Yunfan Xiong, Yupeng Chen, Yuping Lin, Yuqian Cao, Yuqian Wang, Yushun Zhang, Yuting Yan, Yutong Lin, Yuxian Gu, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuxuan Zhou, Yuyang Zhou, Yuzhen Huang, ZF Wu, Zehao Wang, Zehua Zhao, Zehui Ren, Zekai Zhang, Zhangli Sha, Zhe Fu, Zhe Ju, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zheren Gao, Zhewen Hao, Zhibin Gou, Zhicheng Ma, Zhigang Yan, Zhihong Shao, Zhixian Huang, Zhixuan Chen, Zhiyu Wu, Zhizhou Ren, Zhongyu Wu, Zhuoshu Li, Zhuping Zhang, Zian Xu, Zihao Wang, Zihua Qu, Zihui Gu, Zijia Zhu, Zilin Li, Zipeng Zhang, Ziwei Xie, Ziyi Gao, Ziyi Wan, Zizheng Pan, and Zongqing Yao. Deepseek-v4: Towards highly efficient million-token context intelligence, 2026. URL https://arxiv.org/abs/2606.19348.

Jinyang Du, Hao Ma, Xiaohu Shi, Bo Yang, Yanchun Liang, Heow Pueh Lee, and Chunguo Wu. Llm-pdesr: Robust pde discovery via subdomain weighted residuals and llm-guided symbolic hypothesis generation, 2026. URL https://arxiv.org/abs/2607.10546.

Mengge Du, Yuntian Chen, Longfeng Nie, Siyu Lou, and Dongxiao Zhang. Physics-constrained robust learning of open-form partial differential equations from limited and noisy data. *Physics of Fluids*, 36(5), 2024a.

Mengge Du, Yuntian Chen, Zhongzheng Wang, Longfeng Nie, and Dongxiao Zhang. Large language models for automatic equation discovery of nonlinear dynamics. *Physics of Fluids*, 36(9), 2024b.

Mengge Du, Yuntian Chen, and Dongxiao Zhang. Discover: Deep identification of symbolically concise open-form partial differential equations via enhanced reinforcement learning. *Phys. Rev. Res.*, 6:013182, Feb 2024c.

Mingquan Feng, Yixin Huang, Yizhou Liu, Bofang Jiang, and Junchi Yan. PhysPDE: Rethinking PDE Discovery and a Physical Hypothesis Selection Benchmark. In *International Conference on Learning Representations*, volume 2025, pp. 2043–2065, 2025.

Félix-Antoine Fortin, François-Michel De Rainville, Marc-André Gardner, Marc Parizeau, and Christian Gagné. Deap: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, 13(70):2171–2175, 2012.

En-Hao Gao, Cunjing Ge, Yuan Jiang, and Zhi-Hua Zhou. Discovering symbolic partial differential equation by abductive learning. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 57359–57385. Curran Associates, Inc., 2025.

Ting-Ting Gao and Gang Yan. Autonomous inference of complex network dynamics from incomplete and noisy data. *Nature Computational Science*, 2(3):160–168, 2022.

Ting-Ting Gao, Baruch Barzel, and Gang Yan. Learning interpretable dynamics of stochastic complex systems from experimental data. *Nature Communications*, 15(1):6029, 2024.

Sébastien Gaucel, Maarten Keijzer, Evelyne Lutton, and Alberto Tonda. Learning dynamical systems using standard symbolic regression. In *European Conference on Genetic Programming*, pp. 25–36. Springer, 2014.

Gary J Gray, David J Murray-Smith, Yun Li, Ken C Sharman, and Thomas Weinbrenner. Nonlinear model structure identification using genetic programming. *Control Engineering Practice*, 6(11): 1341–1352, 1998.

Arya Grayeli, Atharva Sehgal, Omar Costilla-Reyes, Miles Cranmer, and Swarat Chaudhuri. Symbolic regression with a learned concept library. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 44678–44709, 2024.

Yuchen He, Sung Ha Kang, Wenjing Liao, Hao Liu, and Yingjie Liu. Group projected subspace pursuit for identification of variable coefficient differential equations (gp-ident). *Journal of Computational Physics*, 494:112526, 2023. ISSN 0021-9991.

Zhijian Hu, Yuzhen Wu, Tomas Freire, Erida Gjini, and Kevin Wood. Linking spatial drug heterogeneity to microbial growth dynamics in theory and experiment. *PLOS Computational Biology*, 22 (1):e1013896, 2026.

Dion Häfner, Johannes Gemmrich, and Markus Jochum. Machine-guided discovery of a real-world rogue wave model. *Proceedings of the National Academy of Sciences*, 120(48):e2306275120, 2023.

Subhash Kantamneni, Ziming Liu, and Max Tegmark. OptPDE: Discovering Novel Integrable Systems via AI-Human Collaboration, 2024. URL https://arxiv.org/abs/2405.04484.

Benjamin C. Koenig, Suyong Kim, and Sili Deng. KAN-ODEs: Kolmogorov–Arnold network ordinary differential equations for learning dynamical systems and hidden physics. *Computer Methods in Applied Mechanics and Engineering*, 432:117397, 2024.

Imane Koulali, Erhan Turan, and M. Taner Eskil. LaSIPDE: Latent-Space Identification of Partial Differential Equations from indirect, high-dimensional measurements. *Frontiers in Applied Mathematics and Statistics*, 12:1807939, 2026.

William La Cava, Kourosh Danai, and Lee Spector. Inference of compact nonlinear dynamic models by epigenetic local search. *Engineering Applications of Artificial Intelligence*, 55:292–306, 2016.

Jun Li, Gan Sun, Guoshuai Zhao, and Li-wei H Lehman. Robust low-rank discovery of data-driven partial differential equations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 767–774, 2020.

Qian Li, Yuxiao Hu, Juncheng Liu, and Yuntian Chen. Gensr: Symbolic regression based in equation generative space, 2026. URL https://arxiv.org/abs/2602.20557.

Zeyu Li, Huining Yuan, Wang Han, Yimin Hou, Hongjue Li, Haidong Ding, Zhiguo Jiang, and Lijun Yang. Bi-level identification of governing equations for nonlinear physical systems. *Nature Computational Science*, 5(6):456–466, 2025.

Zichao Long, Yiping Lu, Xianzhong Ma, and Bin Dong. PDE-net: Learning PDEs from data. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 3208–3216. PMLR, 10–15 Jul 2018.

Zichao Long, Yiping Lu, and Bin Dong. Pde-net 2.0: Learning pdes from data with a numeric-symbolic hybrid deep network. *Journal of Computational Physics*, 399:108925, 2019. ISSN 0021-9991.

Siyu Lou, Hao Xu, Wenguan Wang, Lu Lu, Hao Sun, Yang Liu, Linfeng Zhang, Dongxiao Zhang, and Yuntian Chen. Data-driven discovery of governing differential equations across physical systems, 2026. URL https://arxiv.org/abs/2606.09638.

Peter Y Lu, Joan Ariño Bernad, and Marin Soljačić. Discovering sparse interpretable dynamics from partial observations. *Communications Physics*, 5(1):206, 2022.

Yingtao Luo, Qiang Liu, Yuntian Chen, Wenbo Hu, Tian Tian, and Jun Zhu. Physics-guided discovery of highly nonlinear parametric partial differential equations. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, pp. 1595–1607, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701030.

Pingchuan Ma, Tsun-Hsuan Wang, Minghao Guo, Zhiqing Sun, Joshua B. Tenenbaum, Daniela Rus, Chuang Gan, and Wojciech Matusik. LLM and simulation as bilevel optimizers: A new paradigm to advance physical scientific discovery. In *Forty-first International Conference on Machine Learning*, 2024a.

Wenjun Ma, Jun Zhang, Kaikai Feng, Haoyun Xing, and Dongsheng Wen. Dimensional homogeneity constrained gene expression programming for discovering governing equations. *Journal of Fluid Mechanics*, 985:A12, 2024b.

Mikhail Maslyaev, Alexander Hvatov, and Anna Kalyuzhnaya. Data-driven partial derivative equations discovery with evolutionary approach. In *International Conference on Computational Science*, pp. 635–641. Springer, 2019.

Yogesh Chandrakant Mathpati, Tapas Tripura, Rajdip Nayek, and Souvik Chakraborty. Discovering stochastic partial differential equations from limited data using variational bayes inference. *Computer Methods in Applied Mechanics and Engineering*, 418:116512, 2024. ISSN 0045-7825.

Daniel A Messenger and David M Bortz. Weak sindy: Galerkin-based data-driven model selection. *Multiscale Modeling & Simulation*, 19(3):1474–1497, 2021a.

Daniel A. Messenger and David M. Bortz. Weak sindy for partial differential equations. *Journal of Computational Physics*, 443:110525, 2021b. ISSN 0021-9991.

David R. Nelson and Nadav M. Shnerb. Non-hermitian localization and population biology. *Physical Review E*, 58(2):1383–1403, 1998.

Adeel Pervez, Efstratios Gavves, and Francesco Locatello. Mechanistic PDE Networks for Discovery of Governing Equations. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 48962–48973. PMLR, 2025.

Yorgos Psarellis, Seungjoon Lee, Tapomoy Bhattacharjee, Sujit Datta, Juan Bello-Rivas, and Ioannis Kevrekidis. Data-driven discovery of chemotactic migration of bacteria via coordinate-invariant machine learning. *BMC Bioinformatics*, 25, 10 2024.

Chengping Rao, Pu Ren, Yang Liu, and Hao Sun. Discovering nonlinear PDEs from scarce data with physics-encoded learning. In *International Conference on Learning Representations*, 2022.

Patrick AK Reinbold, Logan M Kageorge, Michael F Schatz, and Roman O Grigoriev. Robust learning from noisy, incomplete, high-dimensional experimental data via physically constrained symbolic regression. *Nature communications*, 12(1):3219, 2021.

Kai Ruan, Yilong Xu, Ze-Feng Gao, Yang Liu, Yike Guo, Ji-Rong Wen, and Hao Sun. Discovering physical laws with parallel symbolic enumeration. *Nature Computational Science*, 6:53–66, 2026.

Samuel Rudy, Alessandro Alla, Steven L Brunton, and J Nathan Kutz. Data-driven identification of parametric partial differential equations. *SIAM Journal on Applied Dynamical Systems*, 18(2): 643–660, 2019.

Samuel H. Rudy, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017.

Evgeny S. Saveliev, Samuel Holt, Nabeel Seedat, David L. Bentley, Jim Weatherall, and Mihaela van der Schaar. Influence-guided symbolic regression: Scientific discovery via llm-driven equation search with granular feedback, 2026. URL https://arxiv.org/abs/2605.29184.

Abraham. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 07 1964. ISSN 0003-2700.

Hayden Schaeffer. Learning partial differential equations via data discovery and sparse optimization. *Proceedings. Mathematical, physical, and engineering sciences*, 473(2197):20160446, 2017.

Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *Science*, 324(5923):81–85, 2009.

Parshin Shojaee, Kazem Meidani, Shashank Gupta, Amir Barati Farimani, and Chandan K. Reddy. LLM-SR: Scientific equation discovery via programming with large language models. In *The Thirteenth International Conference on Learning Representations*, 2025.

Robert Stephany and Christopher Earls. Pde-read: Human-readable partial differential equation discovery using deep learning. *Neural Networks*, 154:360–382, 2022.

Robert Stephany and Christopher Earls. Pde-learn: Using deep learning to discover partial differential equations from noisy, limited data. *Neural Networks*, 174:106242, 2024a.

Robert Stephany and Christopher Earls. Weak-pde-learn: A weak form based approach to discovering pdes from noisy, limited data. *Journal of Computational Physics*, 506:112950, 2024b. ISSN 0021-9991.

Fangzheng Sun, Yang Liu, Jian-Xun Wang, and Hao Sun. Symbolic physics learner: Discovering governing equations via monte carlo tree search. In *The Eleventh International Conference on Learning Representations*, 2023.

Rohit Supekar, Boya Song, Alasdair Hastewell, Gary PT Choi, Alexander Mietke, and Jörn Dunkel. Learning hydrodynamic equations for active matter from particle simulations and experiments. *Proceedings of the National Academy of Sciences*, 120(7):e2206994120, 2023.

Cheng Tang, Roy Y. He, and Hao Liu. Wg-ident: Weak group identification of pdes with varying coefficients. *Journal of Computational Physics*, 545:114454, 2026. ISSN 0021-9991.

Pongpisit Thanasutives and Yoshinobu Kawahara. Dynamics-aware identification of governing equations from sparse and noisy data, 2026. URL https://arxiv.org/abs/2607.29036.

Pongpisit Thanasutives, Yoshinobu Kawahara, and Ken-ichi Fukui. Bayesian model selection for variable-coefficient partial differential equation discovery. *Results in Engineering*, 27:106930, 2025.

Pongpisit Thanasutives, Naichang Ke, and Yoshinobu Kawahara. Data-driven sparse identification of governing pdes via knockoff filters and multi-criteria trade-offs, 2026. URL https://arxiv.org/abs/2605.26631.

Silviu-Marian Udrescu and Max Tegmark. Ai feynman: A physics-inspired method for symbolic regression. *Science advances*, 6(16):eaay2631, 2020.

Xingyao Wang, Boxuan Li, Yufan Song, Frank F. Xu, Xiangru Tang, Mingchen Zhuge, Jiayi Pan, Yueqi Song, Bowen Li, Jaskirat Singh, Hoang H. Tran, Fuqiang Li, Ren Ma, Mingzhang Zheng, Bill Qian, Daniel Shao, Niklas Muennighoff, Yizhe Zhang, Binyuan Hui, Junyang Lin, Robert Brennan, Hao Peng, Heng Ji, and Graham Neubig. Openhands: An open platform for AI software developers as generalist agents. In *The Thirteenth International Conference on Learning Representations*, 2025.

Zhenlin Wang, Xun Huan, and Krishna Garikipati. Variational system identification of the partial differential equations governing the physics of pattern-formation: Inference under varying fidelity and noise. *Computer Methods in Applied Mechanics and Engineering*, 356:44–74, 2019.

Zhenlin Wang, Xun Huan, and Krishna Garikipati. Variational system identification of the partial differential equations governing microstructure evolution in materials: Inference over sparse and spatially unrelated data. *Computer Methods in Applied Mechanics and Engineering*, 377:113706, 2021.

Zihang Wei, Yunlong Zhang, Chenxi Liu, and Yang Zhou. Neural networks enabled discovery on the higher-order nonlinear partial differential equation of traffic dynamics, 2025. URL https://arxiv.org/abs/2505.00904.

Shijie Xia, Yuhan Sun, and Pengfei Liu. SR-scientist: Scientific equation discovery with agentic AI. In *The Fourteenth International Conference on Learning Representations*, 2026.

Zikai Xie, Wenmei Li, Man Luo, Jun Jiang, and Linjiang Chen. Language models guide symbolic equation discovery by controlling search, 2026. URL https://arxiv.org/abs/2607.04156.

Haoyun Xing, Jun Zhang, Wenjun Ma, and Dongsheng Wen. Using gene expression programming to discover macroscopic governing equations hidden in the data of molecular simulations. *Physics of Fluids*, 34(5), 2022.

Bing Xiong, Haiyang Fu, Feng Xu, and Yaqiu Jin. Data-driven discovery of partial differential equations for multiple-physics electromagnetic problem, 2019. URL https://arxiv.org/abs/1910.13531.

Hao Xu and Dongxiao Zhang. Robust discovery of partial differential equations in complex situations. *Physical Review Research*, 3(3):033270, 2021.

Hao Xu, Haibin Chang, and Dongxiao Zhang. Dlga-pde: Discovery of pdes with incomplete candidate library via combination of deep learning and genetic algorithm. *Journal of Computational Physics*, 418:109584, 2020. ISSN 0021-9991.

Hao Xu, Dongxiao Zhang, and Junsheng Zeng. Deep-learning of parametric partial differential equations from sparse and noisy data. *Physics of Fluids*, 33:037132, 03 2021.

Hao Xu, Junsheng Zeng, and Dongxiao Zhang. Discovery of partial differential equations from highly noisy and sparse data with physics-informed information criterion. *Research*, 6:0147, 2023.

Hao Xu, Yuntian Chen, Rui Cao, Tianning Tang, Mengge Du, Jian Li, Adrian H Callaghan, and Dongxiao Zhang. Generative discovery of partial differential equations by learning from math handbooks. *Nature Communications*, 16(1):10255, 2025a.

Hao Xu, Yuntian Chen, and Dongxiao Zhang. Beyond empirical models: Discovering new constitutive laws in solids with graph-based equation discovery, 2025b. URL https://arxiv.org/abs/2511.09906.

Hao Xu, Siyu Lou, Yuntian Chen, and Dongxiao Zhang. Joint discovery of governing partial differential equations from multi-source datasets by competitive optimization, 2026. URL https://arxiv.org/abs/2606.30699.

Xiangnan Yu, Hao Xu, Zhiping Mao, Hongguang Sun, Yong Zhang, Zhibo Chen, Dongxiao Zhang, and Yuntian Chen. A data-driven framework for discovering fractional differential equations in complex systems. *Nonlinear Dynamics*, 113(18):24557–24577, 2025.

Ye Yuan, Junlin Li, Liang Li, Frank Jiang, Xiuchuan Tang, Fumin Zhang, Sheng Liu, Jorge Goncalves, Henning U. Voss, Xiuting Li, Jürgen Kurths, and Han Ding. Machine discovery of partial differential equations from spatiotemporal data, 2019. URL https://arxiv.org/abs/1909.06730.

Junsheng Zeng, Hao Xu, Yuntian Chen, and Dongxiao Zhang. Deep learning discovery of macroscopic governing equations for viscous gravity currents from microscopic simulation data. *Computational Geosciences*, 27(6):987–1000, 2023.

Michael Zhang, Samuel Kim, Peter Y. Lu, and Marin Soljačić. Deep learning and symbolic regression for discovering parametric equations. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11):16775–16787, 2024.

Bidan Zhao, Mingming He, and Junwu Wang. Data-driven discovery of the governing equation of granular flow in the homogeneous cooling state using sparse regression. *Physics of Fluids*, 35(1), 2023.

Juncheng Zhong, Chenghuang Shen, Jianfeng Liu, Zhengdong Xiao, Longjiu Luo, Qianrong Wang, Wenjun Xu, and Wenlian Lu. Freeze, then select: Structured field adapters and stability-validated weak selection for pde discovery from sparse observations, 2026. URL https://arxiv.org/abs/2607.29665.

Amirmohammad Ziaei Bideh and Jonathan Gryak. Llm-ode: Data-driven discovery of dynamical systems with large language models. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '26, pp. 853–860. ACM, 2026.

A Complete evaluation protocol

This appendix records the mathematical and experimental details required to interpret the main results. All HEI-based experiments use predictive HEI-v2, fitting and scoring state transitions. The fixed-library baseline instead estimates temporal derivatives for its regression step, as specified in Appendix C.1.

A.1 Candidate language

Candidates use measured column names rather than PDE-specific symbols. A file contains a schema header followed by one complete relation. The recursive grammar is

```
<expression> ::= <atom> | <field> | -<expression> | (<expression>)
               | <expression> <binary-op> <expression>
               | <pointwise-function>(<expression>) | <derivative>
<derivative> ::= D(<expression>, <coordinate-column>)
<field> ::= C<index>(<position-column>[, <position-column> ...])
<binary-op> ::= + | - | * | / | ^
```

The only measured atoms in the formal tasks are two position columns, time, and one scalar reading. Pointwise functions include trigonometric, hyperbolic, exponential, logarithmic, square-root, and absolute-value transforms. The left side must be one derivative of the reading with respect to measurement time; the right side cannot contain a temporal derivative. The formal predictive backend is autonomous and therefore also rejects measurement time used as a right-hand value. Fields may depend on recorded position columns, must be numbered by first appearance, and must enter linearly. A repeated field name preserves one shared function at every tree position. The evaluator accepts or rejects the complete submitted tree and never completes a partial relation.

The language records arbitrary nonempty subsets of the position columns as field signatures. The formal two-dimensional predictive evaluator used here accepts the full signature `C0(sensor_position_1,sensor_position_2)`; lower-dimensional signatures are outside the registered experiment.

Integer tokens have literal semantics: `N0`, `N1`, and `N2` denote 0, 1, and 2. They are not fitted scalar parameters. The formal targets contain only unknown spatial fields, not fitted constants.

The *no algebraic rewrite* instruction in the Agent-facing language refers to candidate identity: HEI does not replace a submitted tree by another hypothesis or add a missing term. Numerical compilation still evaluates the submitted calculus expression, including product-rule lowering and the operator-consistent conservative discretization below. Separately, the post-run structural metric recognizes exact product-rule equivalence; that adjudication never changes the score or selection made during discovery.

A.2 Predictive HEI details

Candidate-faithful lowering. The compiler validates the whole tree and lowers it into

$$\mathcal{F}_\xi[u;\mathcal{C}] = f_0(\mathcal{J}u, \boldsymbol{x}) + \sum_{k=0}^{K-1} \sum_{\alpha} \partial_{\boldsymbol{x}}^{\alpha} c_k(\boldsymbol{x}_{S_k}) f_{k,\alpha}(\mathcal{J}u, \boldsymbol{x}), \tag{8}$$

where $\mathcal{J}u$ is the spatial jet required by the submitted expression. For example, $\partial_x(cu_x)$ induces $cu_{xx} + c_x u_x$ with one shared field, not two free coefficient profiles. Lowering never adds a missing term or changes field identity. The frozen implementation explicitly rejects right-hand-side time dependence and mixed spatial second derivatives; unsupported outcomes remain in the search record and consume a unique-score entry once a canonical candidate has been created.

State preparation. HEI reshapes each raw coordinate table into a $49 \times 49 \times 41$ tensor and applies a degree-three, window-seven Savitzky–Golay filter successively along x and y in every snapshot. At an edge, the filter evaluates the degree-three polynomial fitted to the nearest complete window (`mode=interp`). These prepared states u_{p} supply the initial conditions and targets in Eq. (5). No temporal filtering or temporal derivative estimation is used. Spatial jets are recomputed from the current candidate state at every RK4 stage.

Normalized kernel and continuous field class. The kernel basis approximates each spatial field with a resolution set by anchor density and kernel width, without selecting a symbolic coefficient law. For normalized coordinate $\boldsymbol{z}$ and normalized anchor $\boldsymbol{a}_j$, define $g_j(\boldsymbol{z}) = \exp(-\|\boldsymbol{z} - \boldsymbol{a}_j\|^2/(2\sigma^2))$ and $\kappa_j = g_j / \sum_\ell g_\ell$. For physical coordinate x_q with span $b_q - a_q$, this continuously defined basis satisfies

$$\frac{\partial \kappa_j}{\partial x_q} = \frac{\kappa_j}{\sigma^2 (b_q - a_q)} \left[a_{jq} - \sum_\ell \kappa_\ell a_{\ell q} \right]. \tag{9}$$

Thus one anchor vector defines both a field and its continuous derivatives; field reuse never creates extra coefficients. In the numerical right-hand side, nonconservative field derivatives are centered differences of the constructed grid field, while recognized conservative products are evaluated directly as face fluxes, as specified below.

For each coordinate, the 12 anchor locations are equally spaced empirical quantiles of the 49 unique cell-center coordinates, including their minimum and maximum. Their tensor product gives 144 physical anchor locations. The parameters β_j are kernel weights, not point samples of the field; `evidence` evaluates the resulting function $c(\boldsymbol{x};\boldsymbol{\beta})$ at these anchor locations.

Ridge initialization. At observed states, HEI constructs a linear initialization surrogate in the concatenated anchor vector $\boldsymbol{\beta}$. For nonconservative terms this follows the field-affine lowering directly. For a recognized conservative pair, HEI applies the face-flux operator to each basis column separately; the resulting design initializes, but does not replace, the nonlinear harmonic-face rollout. Integrating this design over each four-step window gives $\boldsymbol{y}_r \approx \boldsymbol{A}_r\boldsymbol{\beta}$ after subtracting field-free terms. Stack the two trajectory systems as $\boldsymbol{y} = [\boldsymbol{y}_A; \boldsymbol{y}_B]$ and $\boldsymbol{A} = [\boldsymbol{A}_A; \boldsymbol{A}_B]$. If s_j is the RMS of column j of $\boldsymbol{A}$, $\boldsymbol{S} = \mathrm{diag}(s_j)$, and $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{S}^{-1}$, the initialization for each kernel width is

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{n}\|\boldsymbol{B}\boldsymbol{\theta} - \boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_2^2, \qquad \widehat{\boldsymbol{\beta}}_0 = \boldsymbol{S}^{-1}\widehat{\boldsymbol{\theta}}, \qquad \lambda = 10^{-8}, \tag{10}$$

where n is the number of stacked transition rows. Numerically zero RMS columns use unit scale. Adam then optimizes the differentiable RK4 prediction loss rather than this linearized objective. We use 120 fit iterations at learning rate 0.035, choose $\sigma \in \{0.04, 0.08, 0.16\}$ by mean development loss, and refit for 180 iterations at learning rate 0.015 on fit plus development transitions.

Proposition A.1 (Shared-field discrimination in the linear initialization)**.** *Let $\boldsymbol{A}_A(h)$ and $\boldsymbol{A}_B(h)$ be the transition designs induced by a candidate h and let $\boldsymbol{y}_A, \boldsymbol{y}_B$ be noiseless transition targets. The linear initialization surrogate has zero joint residual with a shared field vector if and only if*

$$\begin{bmatrix}\boldsymbol{y}_A \\ \boldsymbol{y}_B\end{bmatrix} \in \mathrm{col}\left(\begin{bmatrix}\boldsymbol{A}_A(h) \\ \boldsymbol{A}_B(h)\end{bmatrix}\right). \tag{11}$$

Separate zero-residual fits $\boldsymbol{y}_r \in \mathrm{col}(\boldsymbol{A}_r)$ do not imply this shared condition.

Proof. The stacked vector is in the stated column space exactly when one vector $\boldsymbol{\beta}$ satisfies both linear systems. Separate membership permits different solutions $\boldsymbol{\beta}_A$ and $\boldsymbol{\beta}_B$ and is therefore weaker. □

This proposition describes the initialization geometry, not a global identifiability guarantee for nonlinear rollout. It explains why two excitations constrain compensating fields more strongly than two independent fits.

Operator discretization. First derivatives and nonconservative diagonal second derivatives use centered finite differences in the interior. HEI is not given the simulator boundary condition. It advances every grid value. Along a coordinate q with spacing Δq and indices $0, \ldots, m-1$, the two endpoint rows are

$$(D_q u)_0 = \frac{u_1 - u_0}{\Delta q}, \qquad (D_q u)_{m-1} = \frac{u_{m-1} - u_{m-2}}{\Delta q}, \tag{12}$$

$$(D_{qq} u)_0 = \frac{u_1 - u_0}{\Delta q^2}, \qquad (D_{qq} u)_{m-1} = \frac{u_{m-2} - u_{m-1}}{\Delta q^2}. \tag{13}$$

The first line is one-sided; the second is the cell-centered zero-exterior-flux row. The compiler recognizes the product-rule pair induced by $\partial_q(cu_q)$ and evaluates it as a conservative face flux. Face values use the harmonic mean when adjacent coefficients have the same sign and the arithmetic mean otherwise, with zero flux outside the outer faces. The loss then excludes the outermost cell on each side. This representation distinguishes $c(u_{xx} + u_{yy})$ from $\partial_x(cu_x) + \partial_y(cu_y)$ while preserving a common field across directions.

When one field multiplies diagonal second derivatives with a consistent sign in the submitted tree, optimization parameterizes that field in the corresponding signed softplus chart. Specifically, for the sign $\eta_k \in \{-1, +1\}$ implied by the candidate's principal term, its kernel weights are

$$\beta_{kj} = \eta_k \, \mathrm{softplus}(\gamma_{kj}), \tag{14}$$

where the unconstrained γ_{kj} are initialized from the signed projection of the ridge solution. The chart is selected from the submitted expression and stabilizes a parabolic principal part; it does not use the registered law. Fields attached only to reaction or transport terms remain unconstrained.

Temporal partitions. The 40 recorded intervals are split chronologically into four blocks of 16, 8, 8, and 8 intervals. Nonoverlapping four-step transition starts are constructed inside each block, so no prediction crosses a partition boundary. Thus fit, development, reward, and sealed scoring use four, two, two, and two transition windows per trajectory. The public workspace contains states 0–32. The sealed windows span states 32–36 and 36–40; states 33–40 remain host-only and are used only after submission. The Agent never receives clean states, exact derivatives, the simulator right-hand side, or target fields.

Field-evidence status. For a scored field-bearing candidate, the formal backend returns the label `jointly_fitted_shared_field`. This label states that one coefficient vector was optimized against A and B jointly. It is not a numerical identifiability certificate. The two trial-value arrays in the generic interface are identical copies of that shared fitted field; candidates without fields have no field status.

Table 3: Agent-visible HEI operations. Canonical equations reuse cached scores.

Operation	Returned information
`validate`	Syntax, field signatures, canonical equation, AST size, and validity; no numerical score.
`score`	Candidate ID, A/B and mean reward errors, rank, failure, and remaining budget.
`score-batch`	Independent `score` results for up to 32 complete candidates.
`history`	Archived equations, scores, ranks, and evaluation order.
`evidence`	Fitted fields evaluated at their anchors, dependencies, kernel settings, and field-fit protocol status.
`submit`	Persist one scored candidate, then run sealed evaluation.

B Benchmark construction

B.1 Registered equations and fields

Table 4 gives the exact equations and field transforms used by the generator. Each $z_j(x, y)$ is an independently seeded, zero-mean, unit-standard-deviation draw from the registered Matérn schedule.

Table 4: Generator equations and field transforms.

Family	Equation	Field transform
Variable diffusion	$u_t = D\,(u_{xx} + u_{yy})$	$D = 0.02\exp(0.30z_0)$
Conservative diffusion	$u_t = \partial_x(Du_x) + \partial_y(Du_y)$	$D = 0.02\exp(0.30z_0)$
Reaction–diffusion	$u_t = \partial_x(Du_x) + \partial_y(Du_y) - ku$	$D = 0.02\exp(0.30z_0)$; $k = 1.5\exp(0.50z_1)$
Advection–diffusion	$u_t = D\Delta u - v_x u_x - v_y u_y$	$D = 0.02\exp(0.30z_0)$; $v_x = 0.20z_1$; $v_y = 0.20z_2$
Orthotropic diffusion	$u_t = \partial_x(D_x u_x) + \partial_y(D_y u_y)$	$D_x = 0.04\exp(0.35z_0)$; $D_y = 0.02\exp(0.35z_1)$
Pure advection	$u_t = -v_x u_x - v_y u_y$	$v_x = 0.20z_0$; $v_y = 0.20z_1$
Cubic absorption	$u_t = D\Delta u - ku^3$	$D = 0.02\exp(0.30z_0)$; $k = \exp(0.35z_1)$

B.2 Field morphology schedule

The one-dimensional Matérn-$3/2$ covariance factor is

$$k_\ell(r) = \left(1 + \frac{\sqrt{3}r}{\ell}\right)\exp\left(-\frac{\sqrt{3}r}{\ell}\right). \tag{15}$$

The two-dimensional covariance is the product of its two axis factors. The five environment schedules are:

Environment	Morphology	Length scales / mixture
e01	Smooth	$(0.35, 0.35)$
e02	Medium	$(0.24, 0.24)$
e03	Rough	$(0.16, 0.16)$
e04	Anisotropic	$(0.32, 0.14)$
e05	Multiscale	$0.75(0.36, 0.36) + 0.25(0.12, 0.12)$

Each draw is centered and divided by its grid standard deviation before the transform in Table 4. The covariance jitter is 10^{-10}.

B.3 Excitations, discrete evolution, and noise

All systems use $\Omega = [0,1]^2$ and the cell centers $x_i = y_i = (i + 1/2)/49$, $i = 0, \ldots, 48$. For trajectory $r \in \{A, B\}$, the initial state is

$$u^{(r)}(\boldsymbol{x}, 0) = b_r + a_r \,\mathrm{std}\left[g^{(r)}_{0.30}(\boldsymbol{x}) + 0.05 g^{(r)}_{0.12}(\boldsymbol{x})\right], \tag{16}$$

where the two g fields are independent separable Matérn-$3/2$ draws, std denotes centering and division by the grid standard deviation, $(a_A, a_B) = (0.75, 1.25)$, and $(b_A, b_B) = (-0.5, 0.5)$. The two trajectories share every coefficient field and differ only in these independently drawn excitations.

The generator constructs the discrete operator encoded by each registered law. Nonconservative diffusion applies a cell-centered Neumann Laplacian followed by pointwise multiplication by D. Conservative diffusion uses harmonic coefficient means on interior faces and zero normal flux at the boundary. Advection uses centered first differences in the interior and one-sided first differences at the boundary. The first six systems are linear in the state; their trajectories use SciPy's sparse matrix exponential. Cubic absorption adds the pointwise nonlinear term $-ku^3$ to the nonconservative diffusion operator and uses DOP853, with relative/absolute tolerances $10^{-9}/10^{-11}$ and maximum step 0.01. We record 41 equally spaced states from $t = 0$ to $T = 1$ for the first three families in Table 4, and to $T = 0.75$ for advection–diffusion, orthotropic diffusion, and pure advection, and to $T = 0.5$ for cubic absorption. Neither the boundary rule nor the initial excitations are described to the Agent.

For each trajectory, a standard normal array is centered, standardized, and scaled so that

$$\frac{\mathrm{std}(\varepsilon)}{\mathrm{std}(u_{\mathrm{clean}})} = 0.05. \tag{17}$$

The environment manifest checks finite simulation, integration success, repeat-generation relative error below 10^{-12}, trajectory separation, and the exact requested noise ratio. No environment is selected using Agent or HEI performance.

B.4 Design rationale and freeze boundary

The benchmark varies coefficient morphology independently of equation structure. Environments e01–e03 shorten an isotropic correlation length; e04 introduces different correlation lengths across the two axes; and e05 superposes resolved long- and short-scale components. The shortest registered scale, 0.12, spans approximately 5.9 grid cells. The grid therefore resolves every prescribed covariance scale while leaving the fields nonparametric at the resolution seen by the Agent.

The field transforms encode only qualitative constraints of the corresponding operators. Exponential transforms keep diffusion and reaction rates positive; log scales 0.30–0.50 introduce spatial contrast without isolated singular coefficients. Signed fields with RMS 0.20 allow transport to reverse direction. The 2:1 base ratio in orthotropic diffusion creates directional contrast. The cubic extension uses a positive, independent absorption field and a shorter observation span to resolve nonlinear decay alongside diffusion. The two excitation amplitudes and offsets expose one material realization to different state and derivative combinations, which is required for the shared- field test in Section 3.2. Forty time intervals give the exact $16/8/8/8$ fit/development/reward/sealed partition and admit four-step transitions within every block. The 5% condition is a common, nonzero state- noise operating point applied identically across all families.

These numerical scales were set during development to yield finite, separated, measurably evolving trajectories under one observation budget. Candidate- independent smoothing, field optimization, and transition settings were also selected on the disclosed development audit. The formal benchmark then used disjoint deterministic seeds for coefficient fields, excitations, measurement noise, and evaluator fitting. All generator and HEI settings were frozen before the corresponding formal Agent runs; the cubic extension retains the original Agent and evaluator and uses fresh data-generation seeds. Admission checks test numerical validity and the registered noise/separation conditions only; they never use target rank, structural recovery, or fitted-field accuracy.

B.5 PER-ENVIRONMENT STRUCTURAL OUTCOMES

For a set of N environments, let $c_i \in \{0, 1, 2, 3\}$ be the number of exact successes among the three runs in environment i. We adapt the order-invariant Pass@k estimator of Chen et al. (2021a):

$$\text{Pass@}\, k = \frac{1}{N} \sum_{i=1}^{N} \left[1 - \frac{\binom{3-c_i}{k}}{\binom{3}{k}} \right], \qquad k \in \{1, 2, 3\}. \tag{18}$$

Pass@1 is the order-invariant probability that one randomly selected run succeeds. Because each environment has three runs, Eq. (18) reduces at $k = 1$ to $\sum_i c_i/(3N)$, giving 70 exact runs out of 105 overall. Pass@2 and Pass@3 estimate whether a two- or three-run shortlist contains the registered law. Confidence intervals use 10,000 family-stratified bootstrap samples of complete environments, retaining the three correlated runs in each sampled environment.

The success counts across three Agent runs are shown in Table 5. They are the sufficient statistics for Eq. (18).

Table 5: Exact recoveries among three runs for every frozen environment.

Family	e01	e02	e03	e04	e05
Variable diffusion	3	3	1	3	3
Conservative diffusion	2	3	2	2	0
Reaction–diffusion	3	1	2	2	2
Advection–diffusion	2	2	3	2	2
Orthotropic diffusion	2	1	3	2	1
Pure advection	3	2	0	1	2
Cubic absorption	1	3	2	2	2

Table 6 reports uncertainty for the HER-PDE columns in the main comparison. Bootstrap samples resample complete environments within each family and retain the three runs belonging to each sampled environment.

Table 6: HER-PDE exact structural recovery with 95% environment-bootstrap intervals. Each family contains five environments and 15 Agent runs.

Family	Exact	Pass@1 (%)	Pass@2 (%)	Pass@3 (%)
Variable diffusion	13/15	86.7 [60.0,100]	93.3 [80.0,100]	100 [100,100]
Conservative diffusion	9/15	60.0 [26.7,86.7]	80.0 [40.0,100]	80.0 [40.0,100]
Reaction–diffusion	10/15	66.7 [46.7,86.7]	93.3 [80.0,100]	100 [100,100]
Advection–diffusion	11/15	73.3 [66.7,86.7]	100 [100,100]	100 [100,100]
Orthotropic diffusion	9/15	60.0 [40.0,80.0]	86.7 [73.3,100]	100 [100,100]
Pure advection	8/15	53.3 [20.0,80.0]	73.3 [33.3,100]	80.0 [40.0,100]
Cubic absorption	10/15	66.7 [46.7,86.7]	93.3 [80.0,100]	100 [100,100]
Overall	**70/105**	**66.7 [58.1,75.2]**	**88.6 [80.0,96.2]**	**94.3 [85.7,100]**

B.6 PAIRED OBSERVATION-NOISE CURVE

The noise control replaces the right-hand side of Eq. (17) by $\eta \in \{0.01, 0.05, 0.10\}$ for three representative families: nonconservative variable diffusion, conservative heterogeneous diffusion, and heterogeneous pure advection. These families isolate pointwise field placement, conservative operator scope, and a diffusion-free first-order law. Within each environment, all three levels share the clean trajectory, the centered and standardized Gaussian perturbation, the coefficient and excitation fields, evaluator seeds, Agent, and HEI configuration; only η changes. The 5% row reuses 45 frozen primary runs. The 1% and 10% rows contain 30 newly materialized environment variants and 90 new Agent runs.

One 1% variable-diffusion rollout ended in an infrastructure failure and remains unsuccessful in the registered denominator; the other 14 rollouts in that family recover the exact structure. Three runs exhaust the scientific budget, one at 1% and two at 10%, and count as end-to-end structural failures.

Table 7: Registered outcomes across paired observation-noise levels. Exact entries count successful structures among 15 runs per family and 45 runs overall. Pass@k is computed over the 15 environments at each level.

Metric	1%	5%	10%
Variable diffusion exact	14/15	13/15	13/15
Conservative diffusion exact	11/15	9/15	13/15
Pure advection exact	5/15	8/15	6/15
Overall exact / Pass@1	30/45	30/45	32/45
Pass@2 (%)	77.8	82.2	88.9
Pass@3 (%)	86.7	86.7	93.3

The full curve therefore retains every registered run. Exact recovery does not decrease across the tested range, while pure-advection variation is non-monotone. A 10,000-sample paired bootstrap resamples five complete environments within each family, preserving their three runs and all three noise levels. The resulting Pass@1 intervals are $[53.3, 82.2]$, $[51.1, 82.2]$, and $[60.0, 80.0]$ at 1%, 5%, and 10%, respectively; the paired 10%-minus-1% difference is 4.4 percentage points with interval $[-6.7, 17.8]$.

B.7 Numerical-generator and coefficient-morphology shifts

Independent numerical generator. The solver-shift control reuses the five registered Matérn environments for variable diffusion and conservative diffusion. Realized coefficient fields, A/B excitations, observation-noise arrays, evaluator seeds, observation coordinates, and observation times match the primary benchmark. The shifted generator bilinearly prolongs the field and initial state to a node-centered 99×99 simulation grid, evolves there, and restricts odd interior nodes to the original 49×49 cell centers. Variable diffusion uses an isotropic nine-point Laplacian followed by pointwise coefficient multiplication. Conservative diffusion uses arithmetic coefficient averages at faces; the primary generator and HEI use harmonic averages. All ten scientific runs completed: variable diffusion recovered 5/5 exact structures and conservative diffusion recovered 4/5.

Held-out coefficient morphology. The morphology-shift control keeps the registered equations, positive field transform $D = 0.02 \exp(0.30z)$, simulator, two-excitation protocol, observation grid and times, 5% noise, Agent, and HEI. A disjoint suite seed generates six new environments for each diffusion family. Environments I1–I3 contain five to ten randomly placed, rotated inclusions with radii from 0.04 to 0.18. Environments C1–C3 contain two to five sinuous channel centerlines with half-widths from 0.018 to 0.075. Each latent field is standardized before the positive transform. Admission checks cover deterministic generation, finite trajectories, positive diffusivity, and sharp piecewise morphology; they do not use Agent outcomes or HEI rank.

Table 8: Exact recoveries among three runs for every coefficient-morphology environment. I1–I3 denote discontinuous inclusions and C1–C3 denote sinuous channel networks.

Family	I1	I2	I3	C1	C2	C3	Total
Variable diffusion	3	2	3	3	3	2	16/18
Conservative diffusion	3	3	2	3	3	2	16/18

The morphology suite contains 36 registered runs and 32 exact recoveries. One variable-diffusion I2 run ended in an infrastructure failure and remains an unsuccessful registered run under the no-replacement policy. The two diffusion families contribute 32/36 exact structures and Pass@1/2/3 of $88.9\%/100\%/100\%$. Every environment has two or three exact runs. Both robustness controls focus on diffusion; the primary benchmark separately evaluates reaction and transport structures.

We additionally fix the registered target operator and fit one $c_0(x, y)$ shared across excitations A and B in every morphology-shift environment. This analysis uses the same noisy trajectories, state preparation, 144 Gaussian anchors, and predictive HEI configuration as discovery. Ground-truth fields are interpolated to the anchors only after fitting. Across all 12 fields, the median Pearson

correlation is 0.849 with 95% environment-bootstrap interval $[0.743, 0.869]$, and median relative L_2 error is 0.172 $[0.154, 0.195]$. Table 9 separates the two held-out morphologies.

Table 9: Oracle-structure field recovery under held-out coefficient morphologies. Each morphology cell summarizes three fields per family; Overall summarizes six. Higher Pearson and lower relative L_2 are better.

	Inclusions		Channels	
Family	Pearson	Rel. L_2	Pearson	Rel. L_2
Variable diffusion	0.854	0.151	0.788	0.191
Conservative diffusion	0.842	0.169	0.867	0.176
Overall	0.849	0.154	0.824	0.184

Table 10: Complete family-level structural metrics for the executable controls. Each family contains five environments and 15 runs. Corresponding HER-PDE results and confidence intervals appear in Table 6.

Family	Exact	Pass@1 (%)	Pass@2 (%)	Pass@3 (%)
OpenHands + DeepSeek (no HEI)				
Variable diffusion	1/15	6.7	13.3	20.0
Conservative diffusion	3/15	20.0	33.3	40.0
Reaction–diffusion	0/15	0.0	0.0	0.0
Advection–diffusion	0/15	0.0	0.0	0.0
Orthotropic diffusion	0/15	0.0	0.0	0.0
Pure advection	0/15	0.0	0.0	0.0
Cubic absorption	0/15	0.0	0.0	0.0
Flat Agent + HEI				
Variable diffusion	11/15	73.3	100.0	100.0
Conservative diffusion	2/15	13.3	26.7	40.0
Reaction–diffusion	1/15	6.7	13.3	20.0
Advection–diffusion	3/15	20.0	33.3	40.0
Orthotropic diffusion	4/15	26.7	53.3	80.0
Pure advection	4/15	26.7	46.7	60.0
Cubic absorption	2/15	13.3	26.7	40.0
Recursive typed evolution + HEI				
Variable diffusion	13/15	86.7	93.3	100.0
Conservative diffusion	11/15	73.3	86.7	100.0
Reaction–diffusion	0/15	0.0	0.0	0.0
Advection–diffusion	0/15	0.0	0.0	0.0
Orthotropic diffusion	9/15	60.0	86.7	100.0
Pure advection	6/15	40.0	53.3	60.0
Cubic absorption	0/15	0.0	0.0	0.0
HER-PDE with separate A/B fields				
Variable diffusion	14/15	93.3	100.0	100.0
Conservative diffusion	12/15	80.0	93.3	100.0
Reaction–diffusion	5/15	33.3	40.0	40.0
Advection–diffusion	5/15	33.3	60.0	80.0
Orthotropic diffusion	13/15	86.7	100.0	100.0
Pure advection	6/15	40.0	60.0	80.0
Cubic absorption	7/15	46.7	60.0	60.0

B.8 Field recovery, sealed prediction, and uncertainty

Let $c_f(\boldsymbol{x})$ be a registered field and let $\widehat{c}_j(\boldsymbol{x})$ be a candidate field. Exact structural adjudication supplies a one-to-one mapping $\pi(j) = f$ and the sign $s_j \in \{-1, +1\}$ required by any allowed field-sign reparameterization. The aligned estimate is $\widetilde{c}_f = s_j \widehat{c}_j$. Registered fields are stored on the 49×49 simulator grid and bilinearly interpolated to the 144 physical HEI anchor locations. With the

resulting vectors $\widetilde{\boldsymbol{c}}_f$ and $\boldsymbol{c}_f$, we compute

$$\rho_f = \mathrm{corr}(\widetilde{\boldsymbol{c}}_f, \boldsymbol{c}_f), \qquad e_f = \frac{\|\widetilde{\boldsymbol{c}}_f - \boldsymbol{c}_f\|_2}{\max(\|\boldsymbol{c}_f\|_2, 10^{-12})}. \tag{19}$$

Correlation is undefined, rather than set to zero, if either vector has zero variance. No dense-grid extrapolation enters either metric.

The oracle cohort fits every registered structure once per environment and contains 65 field observations across all 35 environments. The exact-submission field cohort covers all seven families, yielding 129 mapped field observations from 70 exact runs across 33 environments. This cohort weights an environment by its number of exact runs; its median is conditional on successful discovery and is not an environment-matched comparison with the oracle cohort.

Table 11: Family-level field recovery at the 144 HEI anchors. Entries are medians; n counts fitted-field observations. Both fitting cohorts cover all seven families. The two cubic rows separate diffusion and absorption.

	Oracle structure			Exact submissions		
Family	n	Pearson	Rel. L_2	n	Pearson	Rel. L_2
Variable diffusion	5	0.915	0.142	13	0.915	0.142
Conservative diffusion	5	0.955	0.097	9	0.974	0.080
Reaction–diffusion	10	0.936	0.154	20	0.937	0.154
Advection–diffusion	15	0.833	0.496	33	0.833	0.496
Orthotropic diffusion	10	0.759	0.271	18	0.714	0.314
Pure advection	10	0.981	0.194	16	0.988	0.157
Cubic: diffusion D	5	0.852	0.212	10	0.852	0.192
Cubic: absorption k	5	0.403	0.816	10	0.412	0.816
Pooled median	65	0.891	0.232	129	0.878	0.232

The cubic oracle fits use the same noisy trajectories, fit seeds, basis, kernel-width selection, and optimizer as discovery. Field values are taken from scoring before sealed evaluation, with no rescaling against truth. Both fields enter every environment-level comparison, independently of Agent success. Table 12 reports diffusion and absorption estimates separately for each environment.

Table 12: Cubic-absorption oracle fits in all five environments. D and k are the generating fields in $u_t = D\Delta u - ku^3$.

	Diffusion D		Absorption k	
Environment	Pearson	Rel. L_2	Pearson	Rel. L_2
e01	0.852	0.212	0.195	1.341
e02	0.917	0.134	0.421	0.948
e03	0.742	0.279	0.470	0.689
e04	0.853	0.172	0.372	0.612
e05	0.653	0.376	0.403	0.816
Median	0.852	0.212	0.403	0.816

Each sealed block has two nonoverlapping four-interval starts per trajectory. For a submitted run, Eq. (5) is evaluated separately on A and B and the two losses are averaged. Of 105 registered runs, 70 are exact submissions, 34 are submitted non-hits, and one has no submission. The last remains a structural failure and has no sealed prediction datum. The median sealed error is 7.81×10^{-4} across all 104 submissions, 7.11×10^{-4} for exact structures, and 2.01×10^{-3} for non-hits.

The reported structural and field-recovery confidence intervals use 10,000 percentile-bootstrap replicates. The resampling unit is a complete environment: five environments are sampled with replacement inside each family, and all fields and all three Agent runs attached to a sampled environment are retained. This preserves dependence among repeated Agent runs on the same observations. For the 65-field oracle cohort, the pooled median correlation has interval $[0.841, 0.905]$ and relative L_2 error has interval $[0.201, 0.269]$.

Table 13: Fixed-library spatial-field regression over one deterministic fit per environment. TPR and PPV are micro-averages of expanded local support. Field metrics are medians over selected target terms (n_f) against their corresponding expanded coefficient fields.

Family	Exact	TPR (%)	PPV (%)	n_f	Pearson	Rel. L_2
Variable diffusion	0/5	100.0	62.5	10	0.661	0.344
Conservative diffusion	0/5	75.0	100.0	15	0.549	0.412
Reaction–diffusion	0/5	60.0	93.8	15	0.615	0.343
Advection–diffusion	4/5	95.0	86.4	19	0.665	0.453
Orthotropic diffusion	0/5	75.0	100.0	15	0.478	0.398
Pure advection	2/5	100.0	71.4	10	0.964	0.288
Cubic absorption	1/5	93.3	63.6	14	0.409	0.752
Overall	7/35	81.7	81.7	98	0.589	0.400

C Controlled-comparison details

C.1 Fixed-library spatial-field regression

The fixed-library control follows the basis-block, Group Projected Subspace Pursuit (GPSP), and Reduction-in-Residual (RR) design of GP-IDENT (He et al., 2023), with a two-dimensional spatial basis matched to our benchmark. We retain the published RR settings $L = 5$ and $\rho = 0.015$. Two disjoint development environments served as numerical smoke tests; neither recovered its exact expanded target, and their outcomes did not alter the library, filtering, support rule, width set, or fitting protocol. It models

$$u_t(\boldsymbol{x}, t) = \sum_{g \in S} a_g(\boldsymbol{x}) \theta_g(u(\boldsymbol{x}, t)), \qquad a_g(\boldsymbol{x}) = \sum_{m=1}^{144} \beta_{gm} \phi_m(\boldsymbol{x}), \tag{20}$$

where each library term receives an independent normalized Gaussian-basis field. The same global library is supplied to all seven families:

$$\Theta = \{1, u, u^2, u^3\} \cup \{\delta, u\delta, u^2\delta : \delta \in \{u_x, u_y, u_{xx}, u_{xy}, u_{yy}\}\}. \tag{21}$$

Every registered law has an exact product-rule expansion in Θ. The baseline receives neither the family label nor the target support size.

Tensor Savitzky–Golay filters with window length 13 and polynomial order 3 estimate the temporal and spatial derivatives from noisy observations. For each trajectory, the search-visible prefix (states 0–32) is filtered jointly before the resulting derivative rows are partitioned into fit, development, and reward blocks. Temporal stencils cross these internal boundaries, allowing reward observations to contribute to development derivative estimates. Sealed derivative rows are computed in a separate pass; sealed observations do not enter the search-side estimates. For each term, the A/B design blocks are stacked while sharing one coefficient vector $\boldsymbol{\beta}_g$. GPSP constructs support paths for 1–12 active groups. Writing R_k for the fit residual at support size k, RR selects the smallest k satisfying

$$\frac{R_k - R_{k+5}}{5R_1} < 0.015. \tag{22}$$

The development residual selects among Gaussian widths $\{0.04, 0.08, 0.16\}$, after which the chosen support is refitted on the fit and development rows of these preprocessed arrays. Reward residuals are computed after selection; the sealed block is reserved for final evaluation. The expanded support is persisted before the registered support is loaded for adjudication. A predictive HEI replay is performed only after selection and supplies no search-visible score.

The baseline selects all target terms for variable diffusion and pure advection, but additional terms reduce exact recovery to 0/5 and 2/5. It selects three of four target terms in every conservative- and orthotropic-diffusion environment, and three of five in every reaction–diffusion environment. Advection–diffusion yields four of the seven exact expanded supports. All five cubic fits select u^3, but only one selects exactly $\{u^3, u_{xx}, u_{yy}\}$. Together these outcomes produce 98 true-positive local terms, 22 false positives, and 22 false negatives. The field statistics in Table 13 compare

independently fitted expanded coefficients; they do not constitute recovery of the latent material-field identity. Operator nesting and equality or derivative relations among fields are absent from Eq. (20), so compositional recovery and shared-field-identity recovery are N/A. Pass@2 and Pass@3 are also N/A for this deterministic one-fit-per-environment baseline.

All 35 regressions complete. Post-selection predictive HEI evaluation gives finite sealed errors for 30/35 selected equations, with median 7.73×10^{-4} among those 30. The other five comprise four unstable fits and one nonfinite prediction and retain their expanded-support outcomes. Each regression invokes HEI once after selection, for 35 candidate attempts and zero model-provider calls; these are not search feedback.

C.2 Cross-excitation field sharing

The shared-field ablation changes one part of HEI: the identity of a candidate-declared field across the two excitations. Let $\ell_r^P(h, \boldsymbol{\beta})$ be the transition loss for candidate tree h, field anchor vector $\boldsymbol{\beta}$, trajectory $r \in \{A, B\}$, and temporal partition P. The formal evaluator selects one kernel width by mean development loss, refits one anchor vector on the fit and development partitions of both trajectories, and returns

$$E_{\text{shared}}(h) = \frac{1}{2} \sum_{r \in \{A,B\}} \ell_r^{\text{reward}}(h, \widehat{\boldsymbol{\beta}}), \quad \widehat{\boldsymbol{\beta}} \approx \arg\min_{\boldsymbol{\beta}} \sum_{r \in \{A,B\}} \ell_r^{\text{fit+dev}}(h, \boldsymbol{\beta}). \tag{23}$$

The ablation instead selects and refits a field vector separately for each trajectory and returns

$$E_{\text{separate}}(h) = \frac{1}{2} \sum_{r \in \{A,B\}} \ell_r^{\text{reward}}(h, \widehat{\boldsymbol{\beta}}_r), \quad \widehat{\boldsymbol{\beta}}_r \approx \arg\min_{\boldsymbol{\beta}_r} \ell_r^{\text{fit+dev}}(h, \boldsymbol{\beta}_r). \tag{24}$$

Kernel-width selection follows the same distinction: the formal evaluator chooses one width from the mean A/B development loss, whereas the ablation chooses a width for each trajectory. Repeated uses of one field name remain a single field within a trajectory; only its identity across A and B is relaxed. The sealed-final calculation mirrors the corresponding reward calculation. Thus both variants use both excitations, but only Eq. (23) tests whether one material realization can explain them jointly. Proposition A.1 states the corresponding discrimination condition for the linear initialization.

All other components are matched: the 35 registered environments and raw trajectory hashes, three Agent runs per environment, DeepSeek-V4-Flash-0731 settings, Agent prompt and two-phase controller, candidate language, six Agent-visible HEI commands, temporal partitions, Gaussian anchor basis, positive-principal-field chart, optimizer, 96-score ceiling, evaluator-fit seeds, and structural adjudicator. The two configurations differ only in shared versus separate field fitting and within-trajectory scoring. The comparison contains 105 submitted outcomes; cohort accounting is given in Appendix F.

Table 10 gives the complete family results. The separate-field variant recovers 62/105 structures, averages 31.2 distinct HEI evaluations per run, and reaches Pass@1/2/3 of $59.0\%/73.3\%/80.0\%$. Shared fitting recovers 70/105 with $66.7\%/88.6\%/94.3\%$. Its gains concentrate in reaction–diffusion (10 versus 5 exact recoveries), advection–diffusion (11 versus 5), pure advection (8 versus 6), and cubic absorption (10 versus 7). Separate fitting performs better on variable, conservative, and orthotropic diffusion ($14/12/13$ versus $13/9/9$), showing that the invariance constraint exchanges per-trajectory flexibility for stronger discrimination across distinct excitations.

To compare material-field recovery, Table 14 fixes the registered conservative-diffusion tree before fitting, so both variants are evaluated on the same five environments independently of Agent selection. Shared fitting returns one estimate of each registered field. Separate fitting returns two; both A and B estimates enter the truth-recovery medians, with no selection or post-hoc averaging. The A/B column compares the paired trajectory-specific estimates directly.

The two outcomes resolve different parts of the joint task. Separate fitting raises rank-1 structural selection for conservative diffusion by allowing each excitation its own inverse solution. Shared fitting recovers the single registered field more accurately. Across all seven families, its structural advantage appears where reaction and transport operators compete, producing the net gain reported above.

Table 14: Conservative-diffusion structure selection and matched field recovery. Pearson and relative L_2 are medians against the registered field under the fixed registered structure.

Field fit	Exact ↑	Truth Pearson ↑	Truth rel. L_2 ↓	A/B Pearson ↑
Shared (1/env.)	9/15	**0.955**	**0.097**	one field
Separate (2/env.)	**12/15**	0.898	0.143	0.858

C.3 Agent controller and search policy

Flat Agent provides an iterative LLM-plus-numerical-feedback comparison, testing the staged search controller as one bundled intervention. Both variants use the same DeepSeek-V4-Flash-0731 settings, 36 exploration turns, at most two bounded selection turns, complete candidate language, six Agent-visible HEI commands, 96-unique-score ceiling, raw observations, 35 environments, evaluator-fit seeds, and exact structural adjudicator. The Flat Agent removes the turn-16 proposal checkpoint, the requirement to score three distinct hypotheses before the checkpoint, the subsequent raw-data freeze, the dual-lane creative-generation/local-refinement instruction, and periodic score-cadence reminders. It retains bounded final selection from the scored archive. If that archive is empty, the first selection turn allows the Agent to write and score one complete candidate from its accumulated evidence; the remaining turn requires explicit submission. This finalization rule uses the same two-turn selection budget and supplies no candidate structure. Raw arrays remain visible throughout the Flat Agent run.

A valid submission is a parseable, persisted final response that reaches exact structural adjudication. Flat Agent produces 95 valid submissions in 105 runs; the remaining ten exhaust the selection turns without submitting and count as failures. Among the 95 submitted runs, 27 recover the structure (28.4%); the primary metric retains all 105 runs (25.7%). HER-PDE produces 104 submissions. Among matched run identifiers, 19 succeed under both controllers, 51 only under HER-PDE, eight only under Flat Agent, and 27 under neither. The bundled intervention measures whole-controller search and completion under matched resource ceilings; actual evaluation counts depend on the controller. Component-specific effects are not isolated by this comparison.

Typed evolution is our matched search-policy control rather than a reimplementation of DLGA-PDE or SGA-PDE (Xu et al., 2020; Chen et al., 2022). This choice holds the candidate language, coefficient-field model, and numerical objective fixed while removing the language model. DEAP (Fortin et al., 2012) supplies tournament and survivor selection; the scientific grammar and variation operators are implemented in our system. The recursive grammar constructs field-free features of depth at most three from $\{u, u_x, u_y, u_{xx}, u_{yy}\}$, arithmetic, and $\{\sin, \cos, \tanh, \exp, |\cdot|, \log, \sqrt{\cdot}\}$. Multiplication by spatial fields, field reuse/splitting, and flux wrappers construct complete candidates, including $C(x, y)u^3$. Invalid numerical domains are handled by the common evaluator. A target-independent enumeration of shallow coordinate-complete motifs initializes a population of 24. Each generation produces 24 offspring through tournament selection of size 3, crossover with probability 0.45, mutation with probability 0.50, or reproduction. Crossover exchanges feature subtrees or replaces one term; mutation edits features, terms, field identities, or derivative placement. Candidates contain at most five right-hand-side terms. Fitness orders them by HEI error, then term count and tree size; neither structural truth nor sealed error is available during search. Each of three registered seeds per environment stops at 96 distinct HEI scores, 768 proposal attempts, or 64 generations, whichever is reached first. All 105 searches produce valid selected candidates and reach the full score budget, recovering 39 exact structures. The same recursive grammar is used for all seven families, replacing the earlier shallow-grammar experiment in its entirety.

Evolution recovers 24/30 runs across the two single-field diffusion families, but no exact submission for reaction–diffusion, advection–diffusion, or cubic absorption. It expends 10,080 numerical evaluations, compared with 3,103 for HER-PDE, while recovering fewer structures. This comparison measures this registered recursive search policy under the common evaluation budget.

Token telemetry covers all 105 retained runs of each method, including unsuccessful runs. Total tokens are prompt plus completion; cached prompt and reasoning tokens are included in those totals, not added again. Superseded execution attempts are accounted for separately from the retained-run table. Currency is omitted because provider prices vary over execution dates. The distinct HEI totals

Table 15: Submission, evaluation, and provider-token accounting. Token totals include cached prompts and cover the 105 retained runs per method; † marks exhaustion of the 96-evaluation cap in every run. Separate A/B is the field-sharing ablation.

Method	Valid output	Distinct HEI evals/run	Tokens/run (M)	Total (M)
OpenHands + DeepSeek (no HEI)	92/105	N/A	1.157	121.524
Flat Agent + HEI	95/105	13.0	0.962	100.982
Typed evolution + HEI	105/105	96†	0	0
HER-PDE (separate A/B)	105/105	31.2	1.145	120.207
HER-PDE	104/105	29.6	1.136	119.330

are 1,369 for Flat Agent, 10,080 for typed evolution, 3,274 for separate A/B, and 3,103 for HER-PDE. These count numerical candidate evaluations, including failed fits, but exclude cache reads, syntax rejection before scoring, and sealed refits. Internal optimization steps or the two trajectory fits do not count as additional candidate evaluations. OpenHands has N/A because it has no HEI interface; typed evolution makes no provider calls. Versioned configurations and compact result ledgers accompany the code.

Reference HEI latency. We timed one cold, uncached score of the registered structure on environment e01 from each of the original six families. Each measurement ran in a separate process on physical GPU 0, an NVIDIA RTX PRO 5000 72GB Blackwell, with PyTorch 2.7.1, CUDA 12.8, and Python 3.12.13. Timing began after loading both noisy trajectories and constructing the evaluator, and ended after CUDA synchronization. It includes candidate compilation, 120-step fits at each of three registered kernel widths, the selected 180-step refit, and reward-block prediction; it excludes Agent inference, data loading, and sealed-final evaluation.

Table 16: Reference wall time for one uncached HEI score in each of six reference families. $|\mathcal{C}|$ is the number of unknown fields and T is the number of top-level right-hand-side terms. Each entry is one cold e01 measurement.

Family	$\lvert\mathcal{C}\rvert$	T	Wall time (s)
Variable diffusion	1	1	88.9
Conservative diffusion	1	2	160.3
Reaction–diffusion	2	3	174.4
Advection–diffusion	3	3	161.4
Orthotropic diffusion	2	2	123.9
Pure advection	2	2	151.3
Mean	–	–	143.4
Median	–	–	155.8

All six calls returned valid scores. The 88.9–174.4 second range makes the evaluation counts in Tables 2 and 15 concrete while preserving their role as hardware-independent search-efficiency measures.

D Paired success and baseline failure

D.1 Scope of the trace

We report the complete observable scientific record for advection–diffusion environment e03, run 1 under both systems. Completeness here has two parts: every complete equation sent to HEI and every returned score or rejection is listed, and every evidence-driven change of hypothesis family is recorded. Repeated array dumps, shell bookkeeping, and repairs to exploratory scripts are summarized at the operation they support because they do not alter a hypothesis or selection. The HER-PDE record contains 36 model turns, 26 distinct equation files passed to scoring commands, 24 unique canonical evaluation attempts, and one final submission. The paired OpenHands record contains 38 model requests and three final hypotheses; it has no intermediate candidate scorer.

Throughout this section, c_j abbreviates the full spatial-field node $C_j(x, y)$, and $\Delta u = u_{xx} + u_{yy}$. For a valid HEI row, the feedback triple is reward error on A, reward error on B, and their mean. Lower is better.

D.2 HER-PDE: COMPLETE CANDIDATE AND DECISION RECORD

Table 17: Complete chronological HEI record for the exact formal run. Rows I1–I2 were rejected before canonical scoring; all other rows are the 24 unique evaluation attempts in the immutable candidate ledger.

Index	Complete right-hand side $\mathcal{F}[u]$	HEI feedback A/B/mean
1	$c_0\Delta u$	.004007/.004461/.004234
2	$c_0\Delta u + c_1 u + c_2$	.003142/.002871/.003007
3	$c_0 u_{xx} + c_1 u_{yy} + c_2 u_x + c_3 u_y + c_4 u + c_5$	.000568/.000655/.000612
4	$c_0\Delta u + (\partial_x c_0)u_x + (\partial_y c_0)u_y$	.002734/.002609/.002671
5	$\partial_x(c_0 u_x) + \partial_y(c_0 u_y)$	.002734/.002609/.002671
6	$\partial_x(c_0 u_x) + \partial_y(c_1 u_y)$	.002320/.002458/.002389
7	$c_0 u_{xx} + c_1 u_{yy} + c_2 u_x + c_3 u_y + c_4 u$	.000464/.000467/.000466
8	$c_0 u_{xx} + c_1 u_{yy} + c_2 u_x + c_3 u_y$	.000425/.000412/.000419
9	$c_0 u_{xx} + c_1 u_{yy} + c_2 u_x + c_3 u_y + c_4$	.000481/.000450/.000466
10	$\boldsymbol{c_0\Delta u + c_1 u_x + c_2 u_y}$	**.000410/.000394/.000402**
11	$c_0\Delta u + c_1 u_x + c_2 u_y + c_3 u$	.000458/.000449/.000454
12	$c_0 u_{xx} + c_1 u_{yy} + c_2 u$	.003037/.005035/.004036
13	$c_0\Delta u + c_1 u$	.003287/.006204/.004746
14	$0u_{xx} + 0u_{yy} + 1u_x + 2u_y$	.126420/.347092/.236756
15	$0u_{xx} + 0u_{yy} + 1u_x + 2u_y + 3u$	.372904/.519986/.446445
16	$0u_{xx} + 1u_{yy} + 2u_x + 3u_y$	unsupported: no stable shared fit
17	$0u_{xx} + 0u_{yy}$	.003774/.004469/.004121
18	$c_0\Delta u + c_1 u_x + c_1 u_y$	.002184/.007327/.004755
I1	$c_0\Delta u + c_1 u_x + c_2 u_y + c_3 tu$	invalid: explicit time on RHS
I2	$c_0\Delta u + c_1 u_x + c_2 u_y + c_3 t$	invalid: explicit time on RHS
19	$c_0\Delta u + c_1 u_x + c_2 u_y + c_3 u^2$	.000438/.000457/.000448
20	$c_0\Delta u + c_1 u_x + c_2 u_y + c_3 \sin u$	.000455/.000434/.000445
21	$c_0\Delta u + c_1 u_{xy}$	unsupported: mixed derivative
22	$c_0\Delta u + c_1 u_x + c_2 u_y + c_3 u_{xy}$	unsupported: mixed derivative
23	$c_0 u_{xx} + c_1 u_{yy} + c_2 u_{xy}$	unsupported: mixed derivative
24	$c_0 u_{xx} + c_1 u_{yy} + c_2 u_{xy} + c_3 u_x + c_4 u_y$	unsupported: mixed derivative

The integer tokens in rows 14–17 have literal semantics: $N_j = j$; they are shown as $0, 1, 2, 3$ rather than as fitted coefficients. Rows 4 and 5 preserve two different submitted trees—expanded product rule and nested conservative form—although they are mathematically equivalent and receive the same score.

Table 18 reconstructs how this archive changed the Agent's scientific position across the full sequence.

Table 18: Complete evidence-to-decision record for the HER-PDE run.

Stage	Observation or HEI feedback	Scientific update
Raw data	Exact reconstruction gives a $49 \times 49 \times 33$ tensor for each excitation. State ranges contract while spatial patterns persist.	Treat A and B as distinct excitations of one autonomous spatial law; test diffusion, transport, reaction, and source terms.
Derivatives	Direct finite differences give weak constant-library fits; estimated u_t is noisy relative to spatial derivatives.	Use these calculations to generate structures, not to rank a final law; move candidate comparison to predictive HEI.
1–3	Pure variable diffusion scores $.004234$; reaction and source reduce it to $.003007$; a six-field linear law reaches $.000612$.	First derivatives or other omitted linear terms matter, but the six-field law is too broad to identify which ones.
4–7	Expanded and nested conservative diffusion both score $.002671$; orthotropic flux scores $.002389$; adding transport and reaction gives $.000466$.	Reject conservation alone and separate transport from reaction by ablation.
8–13	Removing reaction improves the anisotropic model to $.000419$. Sharing the two diffusion fields gives candidate 10 at $.000402$; adding u, retaining reaction without transport, or adding a source is worse.	Prefer isotropic diffusion plus two independent directional transport fields. This is the first exact registered structure.
Field evidence	Candidate 8 fits diffusion fields with means 0.02026 and 0.02052; candidate 10 uses three fields. Its transport fields have only 0.075 and 0.138 absolute correlation with the corresponding gradient of its diffusion field.	The two second-order channels support field sharing, while the first-order channels should remain independent rather than be forced to ∇c_0.
14–18	Literal coefficients produce errors from $.004121$ to $.446445$ or no stable shared fit; forcing one field for both transport directions gives $.004755$.	Retain spatial fields and distinct x- and y-transport coefficients.
I1–I2	Explicit-time variants are rejected by the autonomous predictive backend.	Do not pursue nonautonomous forcing in this registered search.
19–20	Adding u^2 or $\sin u$ gives $.000448$ and $.000445$.	The extra nonlinear reaction channels are unsupported by predictive improvement.
21–24	Four mixed-derivative constructions receive explicit unsupported results.	Record the backend boundary; do not infer evidence for or against these structures from a missing score.
Selection	Archive history ranks candidate 10 above candidate 8 and every valid augmentation. It uses three fields rather than four or more.	Submit candidate 10. Its sealed-final error is $.000458$; under $c_1 = -v_x$ and $c_2 = -v_y$, it is exactly the registered law.

The target appears at evaluation 10, but discovery does not terminate at the first low score. Evaluations 11–24 test reaction, constant-coefficient, field-sharing, nonlinear, and mixed-derivative alternatives. The final choice therefore follows a falsification sequence: candidate 10 remains both the lowest-error valid hypothesis and the smallest structurally coherent member of its local family. Its A/B search errors are 4.1036×10^{-4} and 3.9375×10^{-4}; its sealed A/B errors are 4.3602×10^{-4} and 4.8042×10^{-4}.

D.3 OpenHands: complete failure decision record

The same raw arrays and run index produce the decision chain in Table 19. OpenHands can write numerical analysis code, but it has neither candidate-specific field fitting nor a predictive hypothesis score. Consequently, none of its intermediate analyses tests the spatial fields declared in its final equations.

Table 19: Complete evidence-to-decision record for the paired OpenHands run.

Stage	Observation	Scientific update
Reconstruction	The two files are reconstructed and round-trip checked as $49 \times 49 \times 33$ arrays with $\Delta x = \Delta y = .02041$ and $\Delta t = .01875$.	Analyze a common local evolution law across A and B.
State summaries	Spatial patterns smooth over time; A and B have different amplitudes and offsets.	Consider diffusion and nonlinear reaction as plausible mechanisms.
Derivative scale	RMS u_t is $1.93/3.14$ for A/B, whereas RMS Δu is $452.9/700.5$.	Expect any diffusion multiplier to be small; construct a broad constant-coefficient library.
Library regression	Separate fits over reaction, advection, and diffusion features obtain $R^2 = .0958$ and $.1509$, with coefficients inconsistent across excitations.	Reject the fitted constants as a reliable shared law.
Derivative audit	Second- versus fourth-order u_t estimates differ by RMS $.446$ against RMS 1.825; temporal second differences exceed first differences.	Attribute much of the regression failure to derivative uncertainty and avoid choosing terms from coefficient magnitude alone.
Correlations	Pairwise correlations of u_t with u, Δu, u^2, u^3, and gradient energy are weak in both trials. Residuals remain spatially patterned.	Infer spatial heterogeneity, but without fitting a spatial field to test which operator carries it.
Final family	Smooth evolution is interpreted as diffusion; spatial residual pattern is interpreted as heterogeneous reaction; response-range asymmetry motivates a quadratic alternative.	Rank heterogeneous cubic reaction–diffusion first, its constant version second, and a quadratic– cubic heterogeneous form third. No advection term is retained.

Table 20: All OpenHands candidates for advection–diffusion e03, run 1. The baseline declares spatial fields but does not fit or evaluate them. All three candidates are structurally incorrect.

Rank	Complete equation
Target	$u_t = C_0(x,y)\Delta u - C_1(x,y)u_x - C_2(x,y)u_y$
1	$u_t = C_0(x,y)\Delta u + C_1(x,y)u + C_2(x,y)u^3 + C_3(x,y)$
2	$u_t = \Delta u + u + u^3 + 1$
3	$u_t = C_0(x,y)\Delta u + C_1(x,y)u + C_2(x,y)u^2 + C_3(x,y)u^3 + C_4(x,y)$

This pair isolates the role of evaluated hypothesis feedback. Both Agents see the same noisy states and both diagnose that direct derivative regression is weak. Only HER-PDE can turn a proposed field-bearing tree into a shared-field prediction test. OpenHands therefore ends with an untested explanation of the spatial residual, whereas HER-PDE uses the residual to generate alternatives and lets their cross-excitation predictions distinguish transport from reaction.

E EXACT HARNESS INPUTS

The following files are copied verbatim from the registered trajectory in Appendix D. They contain generic instrument language and do not name the benchmark family or equation.

E.1 SYSTEM PROMPT

```
You are a scientific hypothesis Agent working in an isolated Bash and
Python environment. Your goal is to infer and submit one complete governing
differential relation that explains both recorded trials.

Begin by reading PROBLEM.md, DSL.md, and data/metadata.json. Inspect the raw NPZ
arrays with Python, NumPy, and SciPy. The experimental record and measured
columns are all the domain context available: do not assume an unreported
physical identity or a known equation family. Decide which numerical
diagnostics and derivatives are informative, propose complete relations under
work/, and use the pde commands to validate and score them. Numerical scoring
is cheap; use it whenever it helps while avoiding semantic duplicates.

Treat cross-trial error as evidence, not an oracle for symbolic similarity or
uniqueness. Compare coherent structural alternatives and penalize unsupported
degrees of freedom in your scientific judgment. You may write arbitrary
analysis scripts under work/. Do not attempt to access files outside the
workspace or use the network.

Before finishing, run pde history --limit 20 --order best and then call
pde submit <candidate_id> for the relation you judge best. A prose answer is
```

```
not a submission. The evaluator never completes a partial relation.
```

E.2 REPRESENTATIVE PROBLEM.MD

```
# Repeated measurements from one instrumented test specimen

An array of probes is mounted across the face of a fixed planar specimen. Every probe
↪  records one continuous scalar response at regular measurement times. The mechanism
↪  controlling the readings is unknown.

## Recorded trials

- Trial A: The specimen was subjected to control program A while all probe readings were
↪  recorded.
- Trial B: The same specimen and probes were used with a different control program B.

The `data/` directory contains the search-visible instrument records. Inspect
the raw arrays freely with Python. Arrays at the same row index form one
observation. Row order carries no physical meaning: reconstruct any tensor grid
from each row's complete coordinate tuple rather than applying a direct reshape
based only on dimension lengths. Determine a complete differential relation
that transfers across both trials. A low numerical error is evidence, not proof
of a unique governing law.
```

E.3 CANDIDATE-LANGUAGE SPECIFICATION

````
# Candidate language

Write one complete differential relation using the recorded column names. A
candidate is a UTF-8 text file with this shape:

```text
field-eqgpt/measured-law@1
<expression> = <expression>
```

The grammar is recursive:

```text
<expression> ::= <atom>
               | <field>
               | -<expression>
               | (<expression>)
               | <expression> <binary-op> <expression>
               | <pointwise-function>(<expression>)
               | <derivative>
<derivative> ::= D(<expression>, <coordinate-column>)
<field> ::= C<index>(<position-column>[, <position-column> ...])
<binary-op> ::= + | - | * | / | ^
```

Recorded columns are `sensor_position_1, sensor_position_2, measurement_time,
↪  sensor_reading`. An atom is a recorded column or
an integer token `N0`, `N1`, and so on. Pointwise functions are `SIN`, `COS`,
`TAN`, `COT`, `SINH`, `COSH`, `TANH`, `COTH`, `EXP`, `LOG`, `SQRT`, `ABS`,
`ARCSIN`, `ARCCOS`, `ARCTAN`, `ARCCOT`, `ARCSINH`, `ARCCOSH`, `ARCTANH`, and
`ARCCOTH`. The second argument of `D` must be a recorded position or time
column. A field's arguments must be distinct recorded position columns.

Every expression alternative in the grammar is valid wherever `<expression>`
appears, including using a `<field>` as the input to `<derivative>`. Before
concluding that a relationship is not representable, write the complete
candidate and run `pde validate`.

The current evaluator accepts exactly one measured time rate on the left and
no time derivative on the right. Number fields `C0`, `C1`, and so on by first
appearance without gaps. Repeating a field name means the same field and
different names mean independently fitted fields. Unknown fields must enter
linearly.

The same field name denotes one fitted field even when its occurrences sit at
different positions in the expression tree, including inside and outside an
operator.

No algebraic rewrite, simplification, expansion, factorization, or term
completion is performed. Every operator remains at the syntax-tree position in
which it was written. Use `pde validate` to check syntax without spending a
````

````
score.

Commands:

```bash
pde validate work/h01.pde
pde score work/h01.pde
pde score-batch work/h01.pde work/h02.pde
pde history --limit 20 --order best
pde evidence <candidate_id>
pde submit <candidate_id>
```

`score-batch` accepts up to 32 complete candidate paths and evaluates each one
independently. The evaluator never adds missing terms. `pde evidence` writes the
↪  already-fitted coefficient fields and numerical diagnostics for one scored candidate to a
↪  JSON file under `data/evidence/`; it does not perform structural analysis.
````

E.4 PHASE-CONTROL PROMPTS

Turn-16 checkpoint.

```
The initial observation phase is over. Before more open-ended analysis, compare a small set
↪  of structurally distinct complete hypotheses and score at least 3 candidates. Your next
↪  Bash action must write the candidate under work/ and end exactly with `pde score
↪  work/<name>.pde`; other actions are rejected until a candidate archive contains 3
↪  distinct scored candidates. They may be tentative: numerical feedback is cheap and is
↪  evidence rather than proof. After the minimum comparison, the run continues in a
↪  candidate-search phase using the complete conversation history; the raw observation
↪  arrays will no longer be mounted.
```

Phase-II search instruction.

```
The initial data-analysis phase is complete. Continue in the same conversation using all
↪  reasoning and tool results above. The raw observation arrays are no longer available in
↪  the sandbox. Candidate files, numerical scoring, archive history, exported fitted-field
↪  evidence, and Bash/Python for analyzing that evidence remain available.

Use two complementary search operations:

- Independent construction: return to the evidence and build a complete candidate without
↪  editing or copying a scored candidate. It must represent a genuinely different
↪  structural explanation; changing only fixed numbers, signs, or a small part of a
↪  previous candidate is not independent construction. Explore the candidate language as
↪  expression trees rather than as a flat catalogue of familiar terms. Cover structural
↪  axes including but not limited to nested and non-nested compositions, grouping and
↪  factor placement across operator boundaries, additive, multiplicative, quotient, and
↪  compositional coupling, shared and independent repeated components, unknown-field
↪  identity and dependencies, operator ordering and interactions, derivative order and
↪  direction, pointwise transformations, and nonlinear combinations. This list is
↪  illustrative: identify additional structural axes from the grammar and the evidence. Do
↪  not assume candidates built from the same primitives are equivalent unless algebra or
↪  the documented evaluator semantics establishes that equivalence. Score a representative
↪  from every materially distinct, evidence-plausible family that fits the budget; do not
↪  blindly take the Cartesian product of all axes. Treat every grammar alternative as a
↪  constructible expression node, including unknown fields. An operator can receive any
↪  expression admitted at that grammar position. Reusing the same unknown-field name at
↪  different tree positions, including inside and outside an operator, preserves one fitted
↪  identity; introducing another field name tests an independent object. Before claiming
↪  that a relationship is inexpressible, write a complete candidate and run `pde validate`.
- Local revision: choose a scored candidate as a parent, state one evidence-based weakness
↪  or ambiguity, make one controlled change, and score the child against the parent.

Before selection, score new complete candidates from both operations. Begin this phase with
↪  independent constructions, then continue using both operations rather than allowing the
↪  current best or most recent lineage to become the template for all proposals. Scores
↪  rank only candidates actually tested; they do not rule out untested structural
↪  explanations. No particular mutation, operator, or equation family is prescribed.
```

Bounded final selection.

```
The exploration phase is over. No new analysis or scoring is allowed. In each Bash tool
↪  call, run exactly one plain command: either `pde history --limit 20 --order best` or `pde
↪  submit <candidate_id>`. Do not use pipes, redirection, `cd`, or combine commands. If you
↪  already know your selection, submit it immediately. Lowest error is not automatically
↪  the true law; weigh physical coherence, unsupported extra terms, and simplicity. You
↪  have at most 2 selection turns.
```

F Reproducibility and baseline completion

Versioned suite configurations record the equations, seed formula, numerical validity checks, Agent budget, and HEI settings. Compact result ledgers map every aggregate entry to a registered rollout. Exact-submission fields come from original scoring exports or replays with the frozen fitting settings. All 62 replayed exact submissions reproduce their recorded sealed errors with maximum absolute difference zero.

The seven-family comparison combines the frozen six-family data with the independently seeded cubic extension. All methods use the same 35 observation pairs. Run replacements preserve the observations and scientific budgets; the result ledger retains both the original and replacement attempts.

For HER-PDE on cubic absorption, the input-contaminated triplets for e03/e05 were excluded and rerun under the frozen protocol; all six rerun outcomes enter the reported cohort.

For Flat Agent, a checkpoint-reactivation repair required 72 reruns and four previously unstarted runs; 29 unaffected outcomes were retained. A subsequent finalization amendment allowed an empty archive to receive one candidate-scoring action during the first selection turn. All 42 affected runs were rerun; the other 63 were retained after replay confirmed identical model inputs, executed commands, and terminal counters under the amended controller. The table includes all 42 new outcomes, irrespective of success, together with those 63 unchanged outcomes.

For separate A/B fields, fourteen infrastructure-invalid executions were rerun. After repairing a missing tool-error reply during final selection, two cubic HTTP-failure runs and one pure-advection selection-exhaustion run were also rerun. All replacement outcomes enter the 105-run cohort.

OpenHands combines 84 original six-family runs, six reruns after provider balance interruptions, and 15 cubic runs. The six interruptions preceded scientific outcomes; each was rerun once under unchanged settings. The cohort contains 92 valid ranked outputs and thirteen invalid finishes counted as failures. Resource totals cover these 105 runs; superseded execution overhead is recorded separately.

G Spatial-drug effective-model case study

G.1 Measurements and discovery inputs

We use arrangement II of the public spatial-drug measurements of Hu et al. (2026). The experiment grows *Enterococcus faecalis* in twelve wells along a row of a 96-well plate, with drug-free wells 3–6 and linezolid in the other eight wells. After each half-hour growth period, an OT-2 pipetting robot exchanges fixed fractions of culture between adjacent wells. At each end it removes the corresponding volume and replenishes it with medium of the local drug condition, implementing population loss to an absorbing exterior. A plate reader measures optical density at 600 nm (OD_{600}) before transfer. The spatial drug conditions set local growth differences; the repeated transfers provide migration. The original study interprets this discrete growth–transfer protocol through a linearized growth–migration PDE, which we use as the reference effective model.

The selected data contain eight technical records at nine times, $t_n = 0.5n$ hours, $n = 0, \dots, 8$. We retain raw OD values, without added noise or spatial smoothing, and use $x_i = i - \frac{1}{2}$, $i = 1, \dots, 12$, in units of well spacing. Discovery inputs A and B average records 1–4 and 5–8, respectively. Thus the same experimental arrangement and inoculation protocol are repeated; these are technical-repeat means, not distinct excitations. The spatial drug profile and independent assay growth rates are absent from the Agent's workspace.

Table 21: All three submissions for spatial-drug arrangement II. Errors are the original HEI scores on the two four-record discovery means, before the post-discovery refits. Search uses transitions 2.5–3.5 h; final uses 3.5–4 h.

Run	Submitted model	Search NMSE	Final NMSE
1	$u_t = \beta u_{xx} + a(x)u - u^2$	0.005900	0.015943
2	$u_t = \beta u_{xx} + s(x) + bu$	0.005952	0.019202
3	$u_t = \beta u_{xx} + a(x)u(1-u)$	0.006388	0.020966

The experimental description supplied to the Agent states:

> A plate reader recorded a scalar microbial-population response along a fixed one-dimensional chain of wells. Local drug exposure is static but unknown, and the same robotic growth-and-transfer protocol was repeated without changing the spatial environment.

It identifies A/B as disjoint four-record means and supplies raw coordinate, time, and measurement arrays. It provides neither the reference equation nor the local drug arrangement. This is discovery with experimental context.

The three conversations retain the two-phase Agent, 16-turn checkpoint, 36-turn budget, two bounded selection turns, and 96-score cap. The one-dimensional numerical instantiation uses one-interval prediction, eight RK4 substeps per half-hour, all twelve observed wells, and the boundary operator below. A fitted scalar is declared with an empty dependency list; a spatial field has signature $\{x\}$. The field basis has twelve anchors and widths $\{0.04, 0.08, 0.16\}$. During discovery, the field optimizer uses 120 Adam fit steps and 180 refit steps, with learning rates 0.035 and 0.015. These experiment-specific settings leave the frozen two-dimensional benchmark unchanged.

G.2 Absorbing boundaries inside the prediction operator

The known exterior loss enters HEI through the discrete forward operator at every RK4 stage, during both field fitting and candidate scoring. For the second derivative, the operator is

$$(L_{\mathrm{abs}}\boldsymbol{u})_i = \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2}, \qquad i = 1, \ldots, 12, \quad u_0 = u_{13} = 0, \quad \Delta x = 1 \text{ well}. \tag{25}$$

The zeros are exterior ghost-site values. The measured endpoints u_1 and u_{12} remain evolving, generally nonzero states and both enter the loss. In particular, the first and last rows are $(u_2 - 2u_1)/\Delta x^2$ and $(u_{11} - 2u_{12})/\Delta x^2$. For a conservative expression $\partial_x(cu_x)$, interior faces use the existing field-averaged flux; the two end rows include exterior losses $-c_1 u_1/\Delta x^2$ and $-c_{12}u_{12}/\Delta x^2$. Constant-field conservative and expanded forms therefore agree. The convention is fixed across candidates, including all three final submissions.

This is a hard boundary condition on the forward map, with no additional boundary penalty and no alteration of the observations. Gradients through the map include its end rows, so the boundary influences the fitted coefficients and the score. No future endpoint measurements are supplied to a prediction.

G.3 All three discovery outcomes

Table 21 records every final submission from the three conversations. Each β is a separately fitted constant and each named spatial coefficient is fitted from the observations. The linear reference has the form $u_t = \beta u_{xx} + a(x)u$; none of the final submissions is algebraically identical to it. These experimental outcomes are assessed as effective models, separately from synthetic exact-structure recovery.

We analyze the first submitted model through matched parameter refitting against the linear reference. Its per-capita reaction rate is $a(x) - \chi u$, with the submitted numeric coefficient corresponding to $\chi = 1\,(\mathrm{OD\,h})^{-1}$. Thus $a(x)$ has units h^{-1} and β has units $\mathrm{well}^2/\mathrm{h}$. In run 2, $s(x)$ has units OD/h and b has units h^{-1}; run 3 implicitly fixes its density scale to 1 OD. Numeric constants are interpreted in the recorded OD and hour units throughout.

Ecological interpretation. Writing the reaction as $a(x)u - \chi u^2 = u[a(x) - \chi u]$ makes its meaning explicit: per-capita growth decreases with density. For $a(x) > 0$ and $\chi > 0$, the local reaction, considered without migration, has a saturation scale $a(x)/\chi$. Heterogeneous logistic growth–diffusion models use this combination of a spatial growth landscape and quadratic density inhibition (Nelson & Shnerb, 1998). Hu et al.'s Supplementary Text, Section 1, starts with a density-dependent Fisher–KPP model and obtains the linear reference by neglecting intraspecies competition (Hu et al., 2026). The discovered quadratic term thus restores a standard finite-density effect to that effective description. Its interpretation concerns density inhibition; the observations do not separate nutrient competition from other microscopic causes of reduced per-capita growth.

The fitted $a(x)$ is the intercept of the nonlinear per-capita rate. Independently assayed linear growth rates therefore do not provide direct ground truth for this coefficient. We report its spatial profile and predictive effect, without assigning a ground-truth field-recovery error. The fixed quadratic was selected for follow-up analysis after inspecting discovery outcomes.

G.4 Matched post-discovery field estimation

Four folds withhold record pairs $(1, 2)$, $(3, 4)$, $(5, 6)$, and $(7, 8)$. Within each fold, the remaining six records, in their original order, form two means of three. Both models fit one shared constant β and one shared spatial field to those two means. Each withheld record is subsequently predicted individually. All eight records participated in the earlier discovery means, so this is post-selection validation with record exclusion during parameter fitting, not a new experiment or an independent assessment of model selection.

The spatial field uses Eq. (4) with twelve anchors along the normalized coordinate $z = (x - x_{\min})/(x_{\max} - x_{\min})$. Normalization affects only basis evaluation; spatial differences use the physical $\Delta x = 1$ well. The twelve Gaussian weights are fitted parameters, whereas the twelve plotted points are the resulting values $a(x_i)$. The linear and fixed-quadratic structures both have thirteen fitted parameters: twelve field weights and one constant diffusion coefficient. The constant is constrained nonnegative; reaction-field weights are unrestricted.

Let $v_{i,n}^{(q)}$ be a training-group mean for group $q \in \{1, 2\}$ and let $\Psi_{\theta,\sigma}$ advance a complete candidate by half an hour under Eq. (25), using eight RK4 substeps. For a set of starting indices $\mathcal{I}$, define $\bar{v}_{q,\mathcal{I}}$ as the mean of $v_{i,n+1}^{(q)}$ over all twelve wells and $n \in \mathcal{I}$. The fitting objective is

$$\mathcal{L}_{\text{drug}}(\mathcal{I}; \theta, \sigma) = \frac{1}{2} \sum_{q=1}^{2} \frac{\dfrac{1}{12|\mathcal{I}|} \displaystyle\sum_{i=1}^{12} \sum_{n \in \mathcal{I}} \left[\Psi_{\theta,\sigma}(\boldsymbol{v}_n^{(q)})_i - v_{i,n+1}^{(q)} \right]^2}{\max\left\{ \dfrac{1}{12|\mathcal{I}|} \displaystyle\sum_{i=1}^{12} \sum_{n \in \mathcal{I}} (v_{i,n+1}^{(q)} - \bar{v}_{q,\mathcal{I}})^2,\ 10^{-12} \right\}}. \tag{26}$$

Every fitting transition starts from its measured group-mean state. The first three transitions (0–1.5 h) fit coefficients for each width; the next two (1.5–2.5 h) select the width. The chosen width is refitted on all five transitions through 2.5 h. Later transitions and withheld records change neither coefficients nor width. Ridge regression with 10^{-8} regularization supplies initialization only; the optimized prediction objective is Eq. (26) without a field or boundary penalty.

For this post-discovery comparison, both structures use float64 trust-region nonlinear least squares with an automatic-differentiation Jacobian through the same RK4 residuals. This replaces the finite-step Adam optimizer used during search, while preserving the Gaussian field space, objective, and boundary operator. To improve conditioning, an unconstrained basis $B = USV^\top$ is reparameterized by $\boldsymbol{w} = VS^{-1}\boldsymbol{z}$, so $B\boldsymbol{w} = U\boldsymbol{z}$; constrained diffusion blocks retain their nonnegative parameter domain. No singular direction is discarded. The three widths all span twelve independent values at the twelve wells; their selection does not identify a physical correlation length in this small-data setting.

We allow 200 function evaluations, use 10^{-10} function, step, and gradient tolerances, and require successful termination with reported first-order optimality at most 10^{-6}. All 48 width-fit/refit solves across the three models below and four folds converge. A first-fold sensitivity check from linear, fixed-quadratic, and earlier negative-quadratic initializations at all three widths converges to the same

Table 22: Post-discovery prediction for arrangement II on eight records withheld in pairs during parameter refitting. Each entry is mean record-level NMSE (lower is better). Persistence retains the measured initial state for the specified horizon.

Model	All one-step	Last one-step	Continuous, 4 h
Linear reference, refitted	0.04493	0.04323	0.09629
Submitted quadratic, $\chi = 1$	**0.04054**	0.03691	**0.08116**
Quadratic, fitted constant χ	0.04138	0.03598	0.08940
Persistence	0.06630	**0.02171**	0.09836

free-quadratic solution. Thus the comparison resolves numerical fitting sensitivity without using withheld prediction errors to choose an initialization.

G.5 Prediction metrics and sensitivity

With fitted parameters frozen, one-step prediction initializes from each record's measured state at every half-hour and scores all eight transitions. Continuous prediction initializes once at 2.5 h and advances to 3, 3.5, and 4 h without measurement resets; its terminal metric scores the twelve wells at 4 h. For each record and comparison, NMSE is the mean squared prediction error divided by the variance of the corresponding observed targets, floored at 10^{-12}. We compute each record's NMSE first, then average the eight values. Thus the mean curves in Figure 3(c) illustrate the spatial prediction, while panel (d) retains record-level errors. Its per-well MAE is computed by taking each record's absolute error before averaging.

The fixed-quadratic model reduces mean one-step NMSE by 9.8% and continuous terminal NMSE by 15.7% relative to the refitted linear reference, with lower error on seven of eight records for both metrics. Persistence is best on the final half-hour interval, where the measured evolution is small. The continuous comparison retains every well, including locations where the linear reference is closer to observations (Figure 4).

Allowing the quadratic coefficient to be fitted adds one parameter. The four folds give $\chi = (0.976, 2.337, 2.274, 2.343)$ in $(\mathrm{OD\,h})^{-1}$, all inhibitory without a sign constraint. The resulting mean one-step and continuous terminal errors improve by 7.9% and 7.2% over the linear reference, respectively. This preserves the aggregate predictive advantage when the submitted numeric coefficient is released, while the fold-to-fold magnitudes show parameter variability. The two other original submissions are retained in Table 21; they were not reranked under this convergence-checked comparison.

The fitted diffusion constant in the fixed-quadratic model ranges from 0.164 to 0.248 well^2/h across folds, and the fitted growth intercepts range from 0.059 to 0.344 h^{-1}. The four field curves in Figure 3(b) show the resulting between-fold variation. Group averaging reduces measurement variation but need not commute with nonlinear evolution: $\overline{u^2} = \bar{u}^2 + \mathrm{Var}(u)$. All compared models use the same training groups, and prediction is assessed on individual records. Technical repeats and overlapping training folds are not treated as independent biological environments for significance claims.

H Scope and failure modes

The frozen synthetic implementation evaluates scalar two-dimensional evolution laws whose unknown fields are time-invariant and enter linearly in the instantaneous right-hand side; state dependence includes the cubic absorption term. It does not yet cover coupled PDE systems, time-varying or stochastic fields, unknown coordinate transformations, or arbitrary boundary operators. The synthetic protocol uses two distinct excitations. The one-dimensional real-data instantiation instead uses technical repeats and known absorbing boundaries as specified in Appendix G.

Two failure modes remain visible in the registered results. First, open-form generation may never propose the target within the finite turn budget. Second, an augmented structure may obtain lower predictive error than the generator on finite noisy observations. Because predictive fit and generating

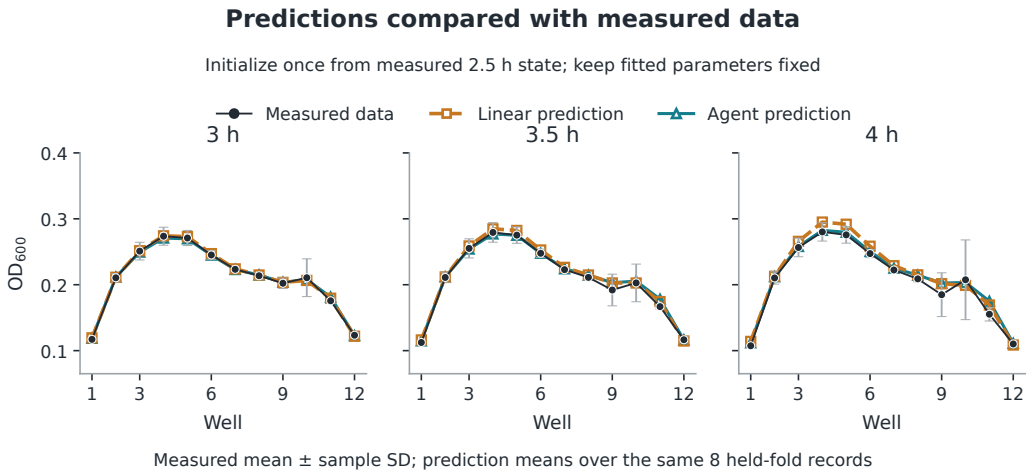


Figure 4: All three continuous forecast times for the spatial-drug case. Black points show measured OD, mean ± sample SD over eight records; orange dashed squares show refitted linear predictions and teal triangles the submitted quadratic model. Both models start each withheld record from its own measured state at 2.5 h and retain fixed coefficients thereafter.

structure can diverge, the Agent weighs field evidence, structural coherence, and unsupported complexity alongside the score. Pure-advection e03 illustrates a remaining selection ambiguity. We therefore report structural recovery, predictive error, and field recovery as separate outcomes.

I Audited Literature-Positioning Pool

Building on the taxonomy and literature survey of Lou et al. (2026), we extend the reviewed set and classify 87 equation-discovery works to construct Figure 1. This is a curated positioning set spanning the method families in Section 2. We classify the most permissive capability demonstrated in each paper's equation-discovery experiments, rather than a capability that might be possible after modifying the method.

The *structure* column uses four categories. *Closed* selects from prespecified terms; *expandable* constructs new terms within a restricted operator or term-generation scheme; *open-form* constructs complete recursive expressions; and *mixed* marks methods whose classification depends on the representation or experiment. The *coefficient* column distinguishes fitted constants, functions constrained to an explicit parametric expression, nonparametric fields, irregular or stochastic fields, and mixed cases. A method occupies the highlighted target cell only if its reported discovery procedure both constructs an open-form operator and estimates candidate-declared nonparametric spatial fields. No work in this curated set meets both criteria.

Table 23: Curated literature-positioning pool (87 publications).

Work / citation	Structure	Coefficient	Note
Equations of motion (Crutchfield & McNamara, 1987)	Mixed	Constant coefficients	Structure depends on representation.
GP structure ID (Gray et al., 1998)	Open-form library	Constant coefficients	
GP-ODE (Cao et al., 2000)	Open-form library	Constant coefficients	
ARE (Bongard & Lipson, 2007)	Open-form library	Constant coefficients	
Free-form laws (Schmidt & Lipson, 2009)	Open-form library	Constant coefficients	
GP dynamics (Gaucel et al., 2014)	Open-form library	Constant coefficients	
ELGP (La Cava et al., 2016)	Open-form library	Constant coefficients	
PDE-FIND (Rudy et al., 2017)	Closed library	Constant coefficients	

Continued on next page

Table 23 (continued)

Work / citation	Structure	Coefficient	Note
SINDy (Brunton et al., 2016)	Closed library	Constant coefficients	
Integral sparse PDE (Schaeffer, 2017)	Closed library	Constant coefficients	
PDE-Net (Long et al., 2018)	Closed library	Expressible by equations	
PDE-Net 2.0 (Long et al., 2019)	Expandable library	Expressible by equations	
DeepMoD (Both et al., 2021)	Closed library	Constant coefficients	
EPDE (Maslyaev et al., 2019)	Expandable library	Constant coefficients	
Multiphysics PDE (Xiong et al., 2019)	Closed library	Constant coefficients	
SGTR (Rudy et al., 2019)	Closed library	Expressible by equations	
VSI pattern formation (Wang et al., 2019)	Closed library	Constant coefficients	
AI Feynman (Udrescu & Tegmark, 2020)	Open-form library	Constant coefficients	
DLGA-PDE (Xu et al., 2020)	Expandable library	Constant coefficients	
Homogenized equations (Arbabi et al., 2020)	Closed library	Constant coefficients	
Low-rank grouped PDE (Li et al., 2020)	Closed library	Mixed	Coefficient depends on representation.
PINN-SR (Chen et al., 2021b)	Closed library	Constant coefficients	
Plasma reduced models (Alves & Fiuza, 2022)	Closed library	Constant coefficients	
Symbolic GNN (Cranmer et al., 2020)	Open-form library	Constant coefficients	
Stepwise-DLGA (Xu et al., 2021)	Expandable library	Expressible by equations	
VSI microstructure (Wang et al., 2021)	Closed library	Constant coefficients	
Weak SINDy (Messenger & Bortz, 2021a)	Closed library	Constant coefficients	
WSINDy-PDE (Messenger & Bortz, 2021b)	Closed library	Constant coefficients	
Active-matter PDE (Supekar et al., 2023)	Closed library	Constant coefficients	
Bayesian VC-PDE (Chen et al., 2026)	Closed library	Mixed	Coefficient depends on representation.
HIN-PDE (Luo et al., 2023)	Closed library	Time-invariant spatial fields	
PCSR (Reinbold et al., 2021)	Closed library	Constant coefficients	
PDE-READ (Stephany & Earls, 2022)	Closed library	Constant coefficients	
R-DLGA (Xu & Zhang, 2021)	Expandable library	Constant coefficients	
SGA-PDE (Chen et al., 2022)	Open-form library	Constant coefficients	
SINDy-AE (Lu et al., 2022)	Open-form library	Constant coefficients	
DHC-GEP (Ma et al., 2024b)	Open-form library	Constant coefficients	
DISCOVER (Du et al., 2024c)	Open-form library	Constant coefficients	
Molecular macro-GEP (Xing et al., 2022)	Open-form library	Constant coefficients	
Network dynamics (Gao & Yan, 2022)	Closed library	Constant coefficients	
PDE-LEARN (Stephany & Earls, 2024a)	Closed library	Constant coefficients	
PeRCN (Rao et al., 2022)	Closed library	Constant coefficients	
PIC (Xu et al., 2023)	Closed library	Constant coefficients	
SEQL / HEQL (Zhang et al., 2024)	Open-form library	Expressible by equations	
SPL (Sun et al., 2023)	Open-form library	Constant coefficients	
GP-IDENT (He et al., 2023)	Closed library	Time-varying / spatiotemporal fields	

Continued on next page

Table 23 (continued)

Work / citation	Structure	Coefficient	Note
Granular PDE (Zhao et al., 2023)	Closed library	Constant coefficients	
Gravity-current PDE (Zeng et al., 2023)	Closed library	Constant coefficients	
ODEFormer (d'Ascoli et al., 2024)	Open-form library	Expressible by equations	
R-DISCOVER (Du et al., 2024a)	Open-form library	Constant coefficients	
Rogue-wave model (Häfner et al., 2023)	Open-form library	Constant coefficients	
Sparse Bayesian PDE (Yuan et al., 2019)	Closed library	Constant coefficients	
SPDE-VB (Mathpati et al., 2024)	Closed library	Other irregular / stochastic fields	
Unknown-equation state estimation (Course & Nair, 2023)	Closed library	Constant coefficients	
Weak-PDE-LEARN (Stephany & Earls, 2024b)	Closed library	Constant coefficients	
E. coli PDE discovery (Psarellis et al., 2024)	Closed library	Constant coefficients	
Fractional PDE (Yu et al., 2025)	Expandable library	Constant coefficients	
KAN-ODEs (Koenig et al., 2024)	Mixed	Constant coefficients	Structure depends on representation.
LaSR (Grayeli et al., 2024)	Open-form library	Expressible by equations	
LLM-SR (Shojaee et al., 2025)	Open-form library	Expressible by equations	
LLM4ED (Du et al., 2024b)	Open-form library	Expressible by equations	
optPDE (Kantamneni et al., 2024)	Expandable library	Constant coefficients	
PhysPDE (Feng et al., 2025)	Closed library	Constant coefficients	
PINN-GP (Cohen et al., 2024)	Open-form library	Constant coefficients	
PSE (Ruan et al., 2026)	Open-form library	Constant coefficients	
Scientific Generative Agent (Ma et al., 2024a)	Open-form library	Expressible by equations	
SiMBA (de Carvalho Servia et al., 2026)	Expandable library	Constant coefficients	
Stochastic dynamics (Gao et al., 2024)	Closed library	Other irregular / stochastic fields	
ABL-PDE (Gao et al., 2025)	Expandable library	Constant coefficients	
Bayesian VC model selection (Thanasutives et al., 2025)	Closed library	Mixed	Coefficient depends on representation.
BILLIE (Li et al., 2025)	Mixed	Constant coefficients	Structure depends on representation.
EqGPT (Xu et al., 2025a)	Open-form library	Expressible by equations	
GraphED (Xu et al., 2025b)	Open-form library	Expressible by equations	
MechNNPDE (Pervez et al., 2025)	Expandable library	Constant coefficients	
TRAFFIC-PDE-LEARN (Wei et al., 2025)	Closed library	Constant coefficients	
WG-IDENT (Tang et al., 2026)	Closed library	Time-invariant spatial fields	
Dynamics-aware ID (Thanasutives & Kawahara, 2026)	Closed library	Constant coefficients	
Equivariant PDE discovery (Ballerin & Grong, 2026)	Closed library	Constant coefficients	
Freeze-then-select (Zhong et al., 2026)	Mixed	Constant coefficients	Structure depends on representation.
GenSR (Li et al., 2026)	Open-form library	Constant coefficients	
IGSR (Saveliev et al., 2026)	Open-form library	Constant coefficients	
KO-PDE-IDENT (Thanasutives et al., 2026)	Closed library	Constant coefficients	
LaSIPDE (Koulali et al., 2026)	Closed library	Constant coefficients	
LLM-ODE (Ziaei Bideh & Gryak, 2026)	Open-form library	Constant coefficients	

Continued on next page

Table 23 (continued)

Work / citation	Structure	Coefficient	Note
LLM-PDESR (Du et al., 2026)	Open-form library	Constant coefficients	
LLM-PySR (Xie et al., 2026)	Open-form library	Constant coefficients	
Multi-source PDE (Xu et al., 2026)	Closed library	Constant coefficients	