Proof or Obstruction: Verifier-Guided Diagnosis of PDE Energy Estimates

Ruiyi Zhang

***Abstract*—Energy estimates for partial differential equations can fail at different levels: an abstract inequality may have no concrete instance, locally valid parameters may be jointly inconsistent, a strict coercivity or dissipation margin may vanish, or the target may lie outside a declared rule library. We formulate a proof-or-obstruction contract that separates these events. The categorical layer records proof composition, abstraction, constrained realization, and the proof-irrelevant Galois polarity; the implementation is a finite ground-Horn shadow with exact symbolic replay. To replace purely synthetic evidence, we extract eight routes from formally published PDE papers. The resulting engineering pre-pilot contains eight source anchors and eight single-change counterfactual failures. Exact plugins replay all anchors and diagnose all mutations, with content-addressed sources and scoped claims. We also retain a 24-task synthetic feedback ablation as a controlled test of the diagnostic interface and run an eight-call open-output route-reconstruction pre-pilot on missing lemmas and energy representations. All eight proposals pass the declared verifier, while exact closed-template baselines solve all four tasks. The current results establish traceable route diagnosis, not a new PDE theorem, a complete account of PDE proof search, or an advantage of a language model over classical symbolic methods.**

***Index Terms*—AI for mathematics, abstract interpretation, Galois polarity, obstruction certificate, PDE energy estimate, verifier-guided reasoning.**

I. Introduction

Let E be a nonnegative energy and let $D = (D_1, \dots, D_d)$ be nonnegative dissipation terms. A typical calculation reaches

$$\frac{1}{2}E'(t) + \lambda \cdot D(t) \le \sum_{\nu=1}^{m} N_\nu(t), \qquad 0 < t < \mathcal{T}, \qquad (1)$$

where $\lambda \in (0,\infty)^d$. The analyst estimates the terms N_ν using Hölder, Sobolev, interpolation, commutator, trace, or problem-specific inequalities, and then chooses Young parameters so that the returned load is strictly smaller than the available dissipation.

The phrase "the estimate does not close" covers several events. An exponent calculation may use a Sobolev endpoint unavailable on the declared domain. Every local inequality may be valid while shared parameters have no joint assignment. A complete chain may exist while a coercivity coefficient vanishes or a Young load reaches its strict budget. The current library may also contain no route to the target. These events require different repairs. A longer search inside the same representation does not tell them apart.

We study a verifier interface with three outcomes. Relative to a finite library of trusted analytic rules, it returns a replayable proof, a replayable obstruction, or `UNKNOWN`. The third outcome is mandatory when search is incomplete or a proposal lies outside the implemented constraint language. In particular, a stable barrier proves only that a target is unreachable in the declared library. It does not prove that the desired analytic estimate is false.

The categorical viewpoint supplies the organizing language. Fully specified estimates are objects, analytic rule instances are multi-input boxes, and finite proof DAGs compose these boxes. A feature functor forgets selected analytic data. Concrete lifts, global constraint sets, and feasible subfibers then separate realization, parameter compatibility, and absorption. Passing to reachability forgets the proof graph and yields a closure operator; its rule-stable truth models form a Galois polarity. Deductive systems and term graphs already have categorical accounts [5], [11], [16]. Our contribution is the PDE-facing diagnostic contract, not a new theory of those categories.

Abstract interpretation organizes sound approximation by closure operators and Galois connections [6], [7]. Counterexample-guided refinement, interpolation, and IC3/PDR use failed reachability attempts to refine an abstraction or construct an inductive certificate [3], [4], [20]. The present framework transfers this separation of proof search and negative certificate to a typed library of PDE estimates. Existing automation has addressed restricted symbolic energy constructions and optimization-based dissipation certificates [1], [12]. Categorical work on differential equations has mainly organized models and solvers [23], [25]; here the object is the estimate proof.

Verifier-guided theorem provers combine learned proposals with formal checking or repair feedback [10], [14], [17], [27]. For PDE estimates, however, exact algebra and finite rule search are strong baselines. A learned component should therefore be tested outside a fixed candidate list: for example, by proposing a missing typed lemma, a new energy term, or a change of trajectory space. The verifier remains responsible for every executable acceptance condition, while a new analytic inequality stays untrusted until independently proved.

This paper makes four contributions.

(i) It states a compositional proof-or-obstruction contract for a finite, acyclic, ground-rule fragment of PDE estimate reasoning. Its negative outputs distinguish missing features, inconsistent parameters, failed strict margins, and rule-relative barriers.

(ii) It introduces a source-backed data envelope containing the published artifact, page and equation locators, a finite route, the trusted primitive boundary, one counterfactual mutation, and exact replay hashes.

Draft v2, 20 July 2026, CST.

(iii) It instantiates the envelope on eight routes from eight published PDE papers. The 16-record engineering pre-pilot contains one proof anchor and one controlled failure per paper. Exact plugins replay all 16 records.

(iv) It places two experiments at the correct claim level. A synthetic closed-repair pilot tests whether certificates affect repair ranking; an open-output pilot tests reconstruction of missing lemma signatures and energy coordinates from four real routes inside declared grammars.

The source-backed set remains a pre-pilot. Formula transcription and rendered pages have been checked, but independent human PDE review and independent encoding review are still pending. None of the counterfactual failures is attributed to the cited authors, and none is used to assert unconditional impossibility for the underlying PDE.

Section II gives the compositional semantics and finite contract. Section III describes the published-route benchmark. Section IV reports the controlled and open-output experiments. Section V states the supported conclusions and remaining obligations.

II. Compositional Semantics and Finite Contract

A. Notation and proof graphs

All vector inequalities are componentwise. An analytic judgment specifies its spatial domain, time interval, function spaces, boundary hypotheses, parameter ranges, and allowed constant dependence. These data remain part of the judgment when a displayed formula suppresses them. Primitive inequalities are trusted only after an analytic proof or an audited source match. The formal kernel checks their typed composition; it does not prove a new Sobolev or interpolation inequality.

Let J be a finite set of fully specified judgments. A primitive rule instance is

$$\rho : (x_1, \dots, x_k) \longrightarrow y, \qquad x_i, y \in J, \tag{2}$$

and R denotes the finite rule library. A rule-labelled proof DAG has an ordered input interface, judgment-labelled wires, and boxes labelled by instances of (2). Sequential substitution is composition, independent calculations form a monoidal product, and fan-out records reuse of an intermediate estimate. Isomorphism classes of these graphs form the free GS-monoidal category $\mathsf{Der}(J, R)$ [5], [11].

Proposition II.1 (Compositional soundness)**.** *If every primitive rule in R is sound for the declared class of admissible PDE objects, every morphism in $\mathsf{Der}(J, R)$ is sound.*

Proof. Order the boxes of the finite DAG topologically. Apply the soundness implication for each box after its incoming wires have been verified. A shared wire retains one concrete judgment at every later use. □

The implementation uses the finite ground-Horn reflection of this category. It retains an enumerated ground rule sequence but does not implement arbitrary term graphs, proof identities, cyclic proofs, or a general resource-enriched category. Table I gives the part of the categorical dictionary used below.

B. Realization, compatibility, and budget

Let q forget selected analytic data while preserving the ordered inputs and outputs of rule instances. For an abstract proof graph p and concrete input judgments $\mathbf{x}$, define

$$\mathrm{Lift}_q(p; \mathbf{x}) = \{\widetilde{p} : q(\widetilde{p}) = p \text{ with inputs } \mathbf{x}\}. \tag{3}$$

An empty set records a concrete type failure, such as a missing mean-zero assumption, an excluded endpoint, insufficient regularity, or a nonuniform constant.

Rule labels in a lift can carry interpolation exponents, Young parameters, and constant policies. A declared constraint Γ_p imposes their global compatibility. We write

$$\mathrm{Lift}_q(p; \mathbf{x} \mid \Gamma_p) = \{\widetilde{p} \in \mathrm{Lift}_q(p; \mathbf{x}) : \widetilde{p} \text{ satisfies } \Gamma_p\}. \tag{4}$$

Every primitive step may have an individual lift while the set in (4) is empty.

For a dissipation budget $\lambda \in (0, \infty)^d$, suppose a compatible lift has load $b_p(\widetilde{p}) \in [0, \infty)^d$ and an admissible remainder. Its feasible subfiber is

$$\mathrm{Feas}_q(p; \mathbf{x} \mid \Gamma_p) = \{\widetilde{p} \in \mathrm{Lift}_q(p; \mathbf{x} \mid \Gamma_p) : b_p(\widetilde{p}) < \lambda, \text{ remainder admissible}\}. \tag{5}$$

Equations (3)–(5) distinguish analytic realization, joint parameter selection, and final closure. These obstructions have a common certificate interface, but they are not all instances of Galois duality: empty lifting fibers are type obstructions, rational infeasibility uses Farkas duality, budget failure uses optimization duality, and only the rule-stable barrier below is the Galois-polarity certificate.

C. Galois shadow and relative barriers

For $S \subseteq J$, let $\mathrm{reach}_R(S)$ be the judgments obtained by finitely many rules in R. A set $B \subseteq J$ is rule-stable when every rule whose premises lie in B also has its conclusion in B. Rule-stable sets are the truth models of the ground proof system. Taking models of a set of judgments and taking the common theory of a set of models form a Galois polarity, and

$$\mathrm{reach}_R(S) = \mathrm{Th}(\mathrm{Mod}(S)). \tag{6}$$

Equation (6) is the proof-irrelevant shadow of the proof category.

Proposition II.2 (Stable-barrier certificate)**.** *If $S \subseteq B$, the set B is rule-stable, and $y \notin B$, then $y \notin \mathrm{reach}_R(S)$.*

Proof. Induction over a finite derivation shows that every reachable judgment lies in B. Hence y has no derivation from S using R. □

The qualifier "using R" is part of the result. A sound new inequality may cross the old barrier. The exact closure $\mathrm{reach}_R(S)$ is always a complete fallback barrier for the finite system, but a compact barrier, such as a derivative-order threshold or an exponent half-space, is more useful as a diagnosis.

TABLE I
PDE ANALYSIS PRACTICE, CATEGORICAL REPRESENTATION, AND EXECUTABLE DIAGNOSIS.

PDE analysis practice	Categorical representation	Verifier-facing object
A fully specified estimate	Object	Typed judgment
Using estimates successively	Composition of rule boxes	Ground proof plan
Reusing an intermediate estimate	Fan-out of one wire	Shared judgment or parameter
Two estimation routes	Parallel morphisms	Several enumerated plans
Keeping only exponents or orders	Feature functor	Abstract plan
Feature plan has no analytic instance	Empty concrete lifting fiber	`LIFT_FAILURE`
Local choices cannot coexist	Empty constrained lifting fiber	`UNSAT_CORE`
Valid choices exceed a strict margin	Empty feasible subfiber	`BUDGET_DUAL`
The library cannot reach the target	Rule-stable truth model	`STABLE_BARRIER`

D. Executable contract

The kernel supports finite ground rules, declared feature requirements, bounded rational linear constraints, scalar linear loads, and finite Horn reachability. Its guarantee stops at that boundary.

Theorem II.3 (Finite proof-or-obstruction contract)**.** *Assume that the primitive rules are sound, the ground signature is finite and acyclic, every admissible abstract plan is enumerated, the declared side conditions are necessary for their rule instances, and each parameter system is a bounded rational polyhedron. The kernel returns one of the following:*

(i) *a* `PROOF` *whose side conditions and shared constraints hold and whose minimum load is strictly below its budget;*

(ii) *a* `STABLE_BARRIER` *when exact Horn closure excludes the target; or*

(iii) *for every enumerated plan, a replayable certificate of missing features, rational infeasibility, or a minimum load at least equal to the strict budget.*

If an operation is unsupported, plan enumeration is incomplete, or a cycle prevents reconstruction, the kernel returns `UNKNOWN`*.*

Proof. Required features are checked before parameter solving. For constraints $Az \le c$, rational Farkas multipliers $\mu \ge 0$ satisfying

$$\mu^{\mathsf{T}} A = 0, \qquad \mu^{\mathsf{T}} c < 0 \tag{7}$$

certify infeasibility. For a feasible plan with load $a^{\mathsf{T}} z$ and scalar budget Λ, a primal feasible point and dual multipliers $v \ge 0$ satisfying

$$A^{\mathsf{T}} v = -a, \qquad -c^{\mathsf{T}} v = L^* \tag{8}$$

certify its minimum. Thus $L^* \ge \Lambda$ excludes the strict budget, while $L^* < \Lambda$ verifies the linear load test. Exact Horn closure and Proposition II.2 verify the relative barrier. The explicit `UNKNOWN` branches prevent a completeness claim outside the stated fragment. □

The current source-backed plugins also check exact nonlinear scalar identities, polynomial signs, and endpoint integrability conditions. These are separate audited plugins, not consequences of the linear kernel theorem. When several abstract plans are reachable, a negative conclusion is supported by the replayable bundle over all enumerated plans. The top-level failure label is only a summary of that bundle.

III. A SOURCE-BACKED PUBLISHED-ROUTE BENCHMARK

A. Unit of data and trusted boundary

The benchmark unit is one finite route used in a formally published paper. Each record contains five layers:

(i) the final journal citation, a SHA-256-addressed PDF, and page, equation, theorem, quote, and formula locators;

(ii) the ordered judgments and estimate steps used in the extracted route;

(iii) the finite list of analytic primitives trusted from the source;

(iv) one declared counterfactual mutation; and

(v) hashes of the original replay, mutated diagnosis, and inverse repair.

The route rather than the paper is the statistical unit. A failed record is labelled as a counterfactual mutation of a successful published route. It is never presented as an error in the source paper.

The verifier checks typed composition, exact algebra, sign conditions, finite-dimensional feasibility, and the declared strict margin. It does not prove the trusted Hardy–Littlewood–Sobolev, Gagliardo–Nirenberg, semigroup, or coercivity estimates from first principles. A proposed missing lemma is therefore marked `PENDING` until it matches a source-verified registry entry or receives an independent analytic proof. Syntactic closure after adding a pending signature is evidence about the route, not evidence that the new inequality is true.

B. Eight published routes

Table II lists the eight source anchors and their single mutations. The displayed margins are exact values returned by the frozen plugins.

The sources cover chemotaxis, nonlinear Schrödinger, generalized MHD, chemotaxis–fluid, damped waves on networks, reaction–diffusion, travelling-wave stability, and Navier–Stokes. Their closure hinges are not all the same kind. They include a free-energy coefficient [2], a sharp interpolation constant [26], a coupled exponent system [28], a transformed entropy sign [19], a modified-energy parameter [9], a polynomial Lyapunov

TABLE II
EIGHT PUBLISHED ESTIMATE ROUTES AND CONTROLLED COUNTERFACTUAL FAILURES. EACH NEGATIVE VERDICT CONCERNS THE DISPLAYED ROUTE ONLY.

Source	Published closure hinge	Single mutation	Exact replay diagnosis
KS-2006 [2]	Logarithmic HLS leaves coefficient $1-\chi M/(8\pi)>0$.	Change $\chi M/\pi$ from 6 to 8.	Coercivity margin changes from $1/4$ to 0; the recorded entropy bound loses its strict coefficient.
NLS-1983 [26]	Sharp Gagliardo–Nirenberg gives $1-(\|\phi_0\|_2/\|\psi\|_2)^{4/N}>0$.	In $N=2$, change the mass ratio from $4/5$ to 1.	Margin changes from $9/25$ to 0; the displayed H^1 coercivity step no longer bounds the gradient.
MHD-2007 [28]	Hölder and two interpolation equations admit witnesses with $\theta_1/(1-\theta_2)\le 2$.	Change (α,β) from $(7/5,11/10)$ to $(13/10,11/10)$.	All interpolation residuals remain zero, while the absorption margin changes from 0 to $-1/11$.
CL-2011 [19]	The signal transform yields quartic coefficient $-\frac{1}{4}(k/\chi)''>0$.	Replace $k/\chi=c/(1+c)$ by $c+c^2$ on $0\le c\le 1$.	The minimum coefficient changes from $1/16$ to $-1/2$; cancellation survives but the recorded quartic dissipation has the wrong sign.
EK-2018 [9]	A cross term with small $\varepsilon>0$ gives an equivalent energy and a strict decay inequality.	With $C_0=C_1=C_P=1$, change ε from $1/8$ to $1/5$.	Energy equivalence still has margin $1/20$, but the dissipation interval is exceeded by $1/30$.
KOU-2001 [15]	Polynomial weights K^{i^2} make every diffusion quadratic form positive semidefinite.	With $(a,b)=(1,4)$, change K from $5/4$ to $6/5$.	The first discriminant changes from 0 to $1764/625>0$; the source sign step is unavailable.
NPR-2008 [24]	The Fourier polynomial $4y^2+(5-\chi)y+1$ is strictly positive on $y\ge 0$.	Change χ from 8 to the threshold 9.	The minimum changes from $7/16$ to 0; strict Fourier coercivity is lost.
LMR-2007 [18]	The direct heat-kernel exponent $1/2+3/(2p)$ is below one for $p>3$.	Change p from 4 to the endpoint 3 in the same unweighted route.	Integrability margin changes from $1/8$ to 0; the source replaces the divergent route by a time-weighted trajectory space.

weight [15], a Fourier coercivity threshold [24], and a time-kernel endpoint [18]. This variety is useful for testing whether one diagnostic interface can preserve distinct mathematical failure modes without collapsing them into "search failed."

C. Extraction and exact replay

The 16 records consist of one `PROOF_ANCHOR` and one `CLOSED_REPAIR` mutation for each paper. Every source artifact hash recomputes; public excerpts contain at most 25 words; quote, formula, input, output, and public-payload hashes are recorded. The records validate against the frozen v0.1 JSON Schema. Applying the inverse mutation restores the source parameters and the proof-anchor verdict.

The deterministic baseline solves 8/8 source anchors and diagnoses 8/8 mutations. This is an engineering result. It shows that the schema, source envelope, exact plugins, and scope discipline work together on eight real routes. It does not show that the kernel can discover the route, prove the trusted analytic primitive, or outperform a language model. In fact, exact symbolic replay is the appropriate baseline for these finite tasks; SymPy and Z3 are used where their supported fragments apply [8], [22].

The current review field remains `PENDING`. The extraction was checked against rendered journal pages, but a public benchmark release requires an independent PDE analyst to rederive every route and a separate encoding auditor to compare each serialized contract with the source. The present paper therefore calls the collection an engineering pre-pilot.

IV. EXPERIMENTS

A. E1: a controlled synthetic diagnostic pilot

E1 asks a narrow question: given a finite list of trusted typed edits, does a replayable obstruction help a fixed reasoning model rank the verifier-accepted repair? The frozen E1-v2 set contains 24 defective PDE-shaped tasks and four accepted sentinels. The defective tasks cross four diagnostic families, three repair mechanisms per family, and two surface variants. Exhaustive candidate replay finds one accepted edit per defective task.

TABLE III
E1 FEEDBACK ABLATION ON 24 SYNTHETIC DEFECTIVE TASKS.

Feedback	Repair@1	Repair@3	MRR
`FAIL`	.333	.542	.417
`LOCAL`	.292	1.000	.590
`FULL`	1.000	1.000	1.000

The fixed model receives one of three diagnoses: a one-bit failure (`FAIL`), the verdict and exact site (`LOCAL`), or the same localization plus the replayable mathematical witness (`FULL`). All other prompt fields are identical. Each isolated call ranks three candidate identifiers. The run used `gpt-5.6-sol` at `xhigh` reasoning effort through Codex CLI 0.144.6 on 20 July 2026. Every task–condition pair used a fresh read-only process without tools, project files, plugins, or web access. There was one generation per pair, and the interface exposed neither a generation seed nor a temperature control. Table III reports the valid run.

All 17 Repair@1 discordances between `FULL` and `LOCAL` favor the full certificate. The exact McNemar p-value is 1.53×10^{-5}, or 4.58×10^{-5} after Holm correction [13], [21]. A certificate-only crossover changes the visible same-site repair while holding the task, candidate order, verdict, and failure site fixed. The model follows the crossed certificate in 24/24 cases.

The E1 witnesses are benchmark-specific finite objects: missing-feature sets, two-interval conflicts, scalar load comparisons, and trusted order frontiers. They should not be read as executions of every general certificate in Theorem II.3. The appropriate negative result is equally important. A deterministic structured-effect router obtains 24/24, and a lexical-overlap policy obtains 22/24 under `FULL`. The pilot therefore supports

typed diagnostic feedback inside a closed repair vocabulary. It provides no evidence that a language model is needed for the task, and it contains no real extracted PDE route. We retain it only as a controlled test of the verifier-to-proposer interface.

B. O2: open-output reconstruction from four real routes

O2 removes the task-specific candidate list. A proposer must return one structured mathematical object drawn from a public output grammar. Four source groups are used:

(i) for KS-2006, the interaction coefficient in $\int_{\mathbb{R}^2} n \log n - a \int_{\mathbb{R}^2} nc$;

(ii) for NLS-1983, the sharp two-dimensional Gagliardo–Nirenberg signature, including its ground-state constant and hypotheses;

(iii) for MHD-2007, the Hölder and interpolation witnesses $(a_1, a_2, \theta_1, \theta_2)$; and

(iv) for CL-2011, a signal transform satisfying $G'(c)^2 = \chi(c)/k(c)$ on the declared interval.

The KS and CL tasks propose an energy coefficient or coordinate; the NLS and MHD tasks propose a missing lemma signature or its exponent realization.

Each task is run in a context-only arm and a typed-obstruction arm. The second condition adds the failure site and the type of missing object, but no coefficient, exponent witness, theorem statement, or finite option list. Fresh processes have no tools, files, web access, or project context. As in E1, the model is `gpt-5.6-sol` at `xhigh` reasoning effort through Codex CLI 0.144.6, with one generation per task and condition. The task-specific JSON schema is fixed across conditions.

Scoring is layered. We record whether the response parses, has the requested type, realizes the route equations, matches trusted evidence, and leaves an open analytic obligation. For KS, MHD, and the declared CL family, acceptance uses exact symbolic identities and inequalities. For NLS, a candidate becomes trusted only by matching the source-verified registry entry with the correct $H^1(\mathbb{R}^2)$ assumptions; algebraic plausibility alone is recorded as an unverified analytic proposal.

All eight runs pass parse, type, route, trust, and final acceptance checks; no run retains an open analytic obligation. The two conditions recover the same mathematical object for every source task:

$$
\begin{aligned}
a &= \frac{3}{4}, \\
\|u\|_{L^4(\mathbb{R}^2)}^4 &\le \frac{2}{\|Q\|_{L^2(\mathbb{R}^2)}^2} \|u\|_{L^2(\mathbb{R}^2)}^2 \|\nabla u\|_{L^2(\mathbb{R}^2)}^2, \\
(a_1, a_2, \theta_1, \theta_2) &= \left(\frac{30}{11}, \frac{60}{19}, 1, \frac{1}{2} \right), \\
G(c) &= \sqrt{c(1+c)} + \operatorname{arsinh}(\sqrt{c}) + C.
\end{aligned}
$$

The first value uses the declared coefficient $\chi = 3/2$. The NLS output is trusted through the source registry rather than an independent proof of the sharp inequality.

Mean output tokens are 1327.5 under `CONTEXT_ONLY` and 353.25 under `TYPED_OBSTRUCTION`; mean latencies are 38.84 and 14.55 seconds. These aggregates are dominated by the MHD context-only run, which uses 4537 output tokens and 112.75 seconds. There is one run per task and condition, and the remaining three task pairs do not show a uniform resource contrast. The resource values are descriptive and do not establish an efficiency effect. Since both conditions obtain 4/4, the pre-pilot also supplies no efficacy contrast.

These tasks remain source-route reconstruction. An accepted output recovers one declared representation or instantiates a published analytic signature. It does not prove a new inequality or PDE theorem. Exact closed-template baselines solve 4/4: rational identity solving for KS, registry lookup for NLS, exact rational feasibility for MHD, and coefficient matching in the supplied CL transform grammar. These baselines expose the ceiling of the present task. They do not invent the free-energy template, the analytic inequality family, the interpolation ansatz, or the transform grammar. With one call per task and condition, the experiment is a pipeline and failure-mode pre-pilot, not a statistical comparison of prompt conditions.

C. What the experiments test

The three empirical components have different functions.

- The eight-paper exact replay tests source provenance, executable contracts, single-mutation diagnosis, and claim scope.
- E1 tests whether a diagnostic witness can change repair ranking when a finite trusted vocabulary is already supplied.
- O2 tests whether a proposer can construct a typed object without a task-specific option list and whether the deterministic verifier can separate route realization from analytic trust.

None tests unrestricted discovery of a new PDE energy method. In particular, success on a fixed coefficient grammar or a finite rational system can be matched by classical enumeration or exact solving. The next benchmark must hide an energy atom or request a lemma outside a source-verified registry, then require independent proof or blinded PDE review before promotion to the trusted library.

V. Discussion and Conclusion

A. Supported conclusions

The categorical view is useful because it keeps several questions separate. The proof category records how estimates compose and where intermediate judgments are shared. Feature abstraction records the analyst's exponent or derivative bookkeeping. Lifting asks whether that bookkeeping has an analytic instance. Global constraints ask whether local choices coexist. A budget test asks whether the realized route closes. The Galois shadow asks what is reachable after proof identity has been forgotten.

The source-backed pre-pilot shows that this interface can encode and replay eight mathematically different closure hinges from eight published papers. The plugins recover every source anchor and distinguish a vanished free-energy coefficient, sharp-mass coercivity, an absorption inequality, a transformed-entropy sign, a modified-energy interval, a quadratic-form sign, a Fourier threshold, and a divergent endpoint kernel. This is evidence for the coverage of the diagnostic vocabulary, subject

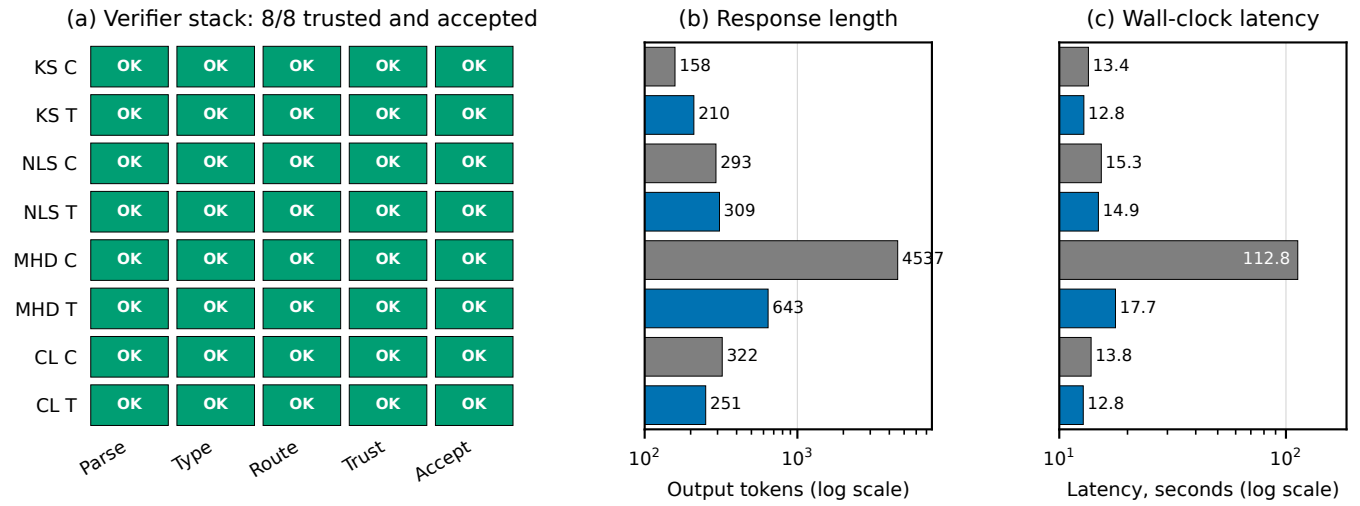


Fig. 1. O2 verifier outcomes and descriptive resource measurements. C is context only and T is typed obstruction. There is one run per task–condition pair; all acceptance endpoints are at ceiling, and the MHD context-only call dominates the aggregate token and latency contrast.

to the pending independent review. It is not evidence for completeness over PDE analysis.

E1 shows that certificate content can determine a repair ranking inside a closed candidate set, but the deterministic controls deny an LLM-specific claim. O2 removes the finite task-specific list but retains a declared output grammar and exposes the trust boundary. Its accepted objects still reconstruct known routes, so the result should be read as verifier plumbing and structured proposal testing, not unrestricted discovery.

B. What the results do not support

Four conclusions are excluded by the design.

(i) A counterfactual route obstruction is not a theorem that the PDE has no solution or no alternative estimate.

(ii) A stable barrier is relative to the frozen rule library. Adding a sound new inequality can cross it.

(iii) A route that closes after adding a proposed lemma does not prove that lemma. The proposal remains `PENDING` until independently verified.

(iv) Closed-world success does not establish an AI advantage. Exact search, symbolic algebra, SMT, and convex optimization are the natural baselines.

The implementation is also smaller than the categorical semantics. It is a finite ground-Horn shadow, not a proof assistant for PDEs. It does not yet represent arbitrary shared proof DAGs, quantify over function spaces, prove boundary identities, or synthesize a complete family of compact barriers. Unsupported cases return `UNKNOWN`.

C. Next experiments

The next useful task is proof-system expansion. A model may propose a new intermediate quantity, energy atom, variable transform, test function, or typed inequality. Classical tools then enumerate and verify consequences in the expanded system. A failed attempt should return a typed obstruction that guides the next proposal.

Four real routes already indicate suitable directions. Egger–Kugler's ordinary energy must be enlarged by a cross term to obtain exponential decay [9]. Kouachi's coefficient recurrence determines the polynomial weights K^{i^2} [15]. Nadin–Perthame–Ryzhik use a specific one-dimensional interpolation inequality to close their nonlinear estimate [24]. Lemarié-Rieusset repairs a critical endpoint by changing the time-weighted trajectory representation [18]. These examples allow a graded benchmark: coefficient recovery inside a fixed basis, discovery of an atom from a public grammar, and genuinely open analytic proposals requiring human or formal review.

Source splits must be by paper rather than by mutation. Repeated generations, classical grammar-search baselines, type-swapped crossover controls, and blinded review are needed before comparing proposal systems. A compact dual certificate is a second open task: instead of searching only for a proof, a system may propose an exponent potential, derivative count, scaling law, or budget lower bound and have its stability checked on the primitive rules.

D. Conclusion

A useful PDE estimate verifier should return more than success or timeout. It should identify whether an abstract route lacks an analytic instance, has incompatible parameters, loses a strict closure margin, or is blocked by its current rule library. The eight published routes show that these diagnoses occur in ordinary PDE analysis and can be represented by one auditable data contract. The synthetic pilot shows that the certificates can guide a proposer, while its classical controls show that finite repair ranking is not the research problem. The next problem is to propose a change of proof language and then earn its admission through deterministic verification and independent analytic review.

Data and code availability

The implementation, tests, source envelopes, audit notes, prompts, raw model events, and valid artifacts accompany this manuscript under `prototypes/pde_closure`. The published PDFs used for the internal source audit are content-addressed in each task record. Redistribution of source articles remains subject to the publishers' terms. The 16-record collection is an engineering pre-pilot and is not labelled as an independently reviewed benchmark release.

Conflict of interest

The author declares no conflict of interest.

Declaration on generative AI

Ruiyi Zhang proposed the connection between PDE energy-estimate closure and Galois duality and set the research question. OpenAI's GPT-5.6 Sol, accessed through Codex, assisted with the categorical implementation, prototype code, source-route extraction, experimental design and execution, and drafting. The same model family was evaluated in the isolated proposal experiments; those processes had no access to project files, tools, or source answers. The model was not treated as a source for mathematical or bibliographic claims. Responsibility for the statements and citations remains with the author.

REFERENCES

[1] M. Ahmadi, G. Valmorbida, and A. Papachristodoulou, "Dissipation inequalities for the analysis of a class of PDEs," *Automatica*, vol. 66, pp. 163–171, 2016, doi: 10.1016/j.automatica.2015.12.010.

[2] A. Blanchet, J. Dolbeault, and B. Perthame, "Two-dimensional Keller–Segel model: Optimal critical mass and qualitative properties of the solutions," *Electron. J. Differential Equations*, vol. 2006, no. 44, pp. 1–33, 2006.

[3] A. R. Bradley, "SAT-based model checking without unrolling," in *Verification, Model Checking, and Abstract Interpretation*, LNCS 6538. Berlin, Germany: Springer, 2011, pp. 70–87, doi: 10.1007/978-3-642-18275-4_7.

[4] E. M. Clarke, O. Grumberg, S. Jha, Y. Lu, and H. Veith, "Counterexample-guided abstraction refinement for symbolic model checking," *J. ACM*, vol. 50, no. 5, pp. 752–794, 2003, doi: 10.1145/876638.876643.

[5] A. Corradini and F. Gadducci, "An algebraic presentation of term graphs, via GS-monoidal categories," *Appl. Categorical Struct.*, vol. 7, no. 4, pp. 299–331, 1999, doi: 10.1023/A:1008647417502.

[6] P. Cousot and R. Cousot, "Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation of fixpoints," in *Proc. 4th ACM SIGPLAN–SIGACT Symp. Principles of Programming Languages*, 1977, pp. 238–252, doi: 10.1145/512950.512973.

[7] P. Cousot and R. Cousot, "Systematic design of program analysis frameworks," in *Proc. 6th ACM SIGPLAN–SIGACT Symp. Principles of Programming Languages*, 1979, pp. 269–282, doi: 10.1145/567752.567778.

[8] L. de Moura and N. Bjørner, "Z3: An efficient SMT solver," in *Tools and Algorithms for the Construction and Analysis of Systems*, LNCS 4963. Berlin, Germany: Springer, 2008, pp. 337–340, doi: 10.1007/978-3-540-78800-3_24.

[9] H. Egger and T. Kugler, "Damped wave systems on networks: Exponential stability and uniform approximations," *Numer. Math.*, vol. 138, pp. 839–867, 2018, doi: 10.1007/s00211-017-0924-4.

[10] E. First, M. N. Rabe, T. Ringer, and Y. Brun, "Baldur: Whole-proof generation and repair with large language models," in *Proc. ACM Joint Eur. Software Eng. Conf. Symp. Foundations of Software Engineering*, 2023, pp. 1229–1241, doi: 10.1145/3611643.3616243.

[11] T. Fritz and W. Liang, "Free GS-monoidal categories and free Markov categories," *Appl. Categorical Struct.*, vol. 31, art. 21, 2023, doi: 10.1007/s10485-023-09717-0.

[12] T. Hagstrom and D. Appelö, "Automatic symmetrization and energy estimates using local operators for partial differential equations," *Commun. Partial Differential Equations*, vol. 32, no. 7, pp. 1129–1145, 2007, doi: 10.1080/03605300600854258.

[13] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Stat.*, vol. 6, no. 2, pp. 65–70, 1979.

[14] T. Hubert *et al.*, "Olympiad-level formal mathematical reasoning with reinforcement learning," *Nature*, vol. 651, pp. 607–613, 2026, doi: 10.1038/s41586-025-09833-y.

[15] S. Kouachi, "Existence of global solutions to reaction–diffusion systems via a Lyapunov functional," *Electron. J. Differential Equations*, vol. 2001, no. 68, pp. 1–10, 2001.

[16] J. Lambek, "Deductive systems and categories I: Syntactic calculus and residuated categories," *Math. Syst. Theory*, vol. 2, no. 4, pp. 287–318, 1968, doi: 10.1007/BF01703261.

[17] G. Lample, M.-A. Lachaux, T. Lavril, X. Martinet, A. Hayat, G. Ebner, A. Rodriguez, and T. Lacroix, "HyperTree proof search for neural theorem proving," in *Advances in Neural Information Processing Systems*, vol. 35, 2022.

[18] P. G. Lemarié-Rieusset, "The Navier–Stokes equations in the critical Morrey–Campanato space," *Rev. Mat. Iberoam.*, vol. 23, no. 3, pp. 897–930, 2007, doi: 10.4171/RMI/518.

[19] J.-G. Liu and A. Lorz, "A coupled chemotaxis–fluid model: Global existence," *Ann. Inst. H. Poincaré C Anal. Non Linéaire*, vol. 28, no. 5, pp. 643–652, 2011, doi: 10.1016/j.anihpc.2011.04.005.

[20] K. L. McMillan, "Interpolation and SAT-based model checking," in *Computer Aided Verification*, LNCS 2725. Berlin, Germany: Springer, 2003, pp. 1–13, doi: 10.1007/978-3-540-45069-6_1.

[21] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947, doi: 10.1007/BF02295996.

[22] A. Meurer *et al.*, "SymPy: Symbolic computing in Python," *PeerJ Comput. Sci.*, vol. 3, art. e103, 2017, doi: 10.7717/peerj-cs.103.

[23] L. Morris, A. Baas, J. Arias, M. Gatlin, E. Patterson, and J. P. Fairbanks, "Decapodes: A diagrammatic tool for representing, composing, and computing spatialized partial differential equations," *J. Comput. Sci.*, vol. 81, art. 102345, 2024, doi: 10.1016/j.jocs.2024.102345.

[24] G. Nadin, B. Perthame, and L. Ryzhik, "Traveling waves for the Keller–Segel system with Fisher birth terms," *Interfaces Free Bound.*, vol. 10, no. 4, pp. 517–538, 2008, doi: 10.4171/IFB/200.

[25] E. Patterson, A. Baas, T. Hosgood, and J. Fairbanks, "A diagrammatic view of differential equations in physics," *Math. Eng.*, vol. 5, no. 2, pp. 1–59, 2023, doi: 10.3934/mine.2023036.

[26] M. I. Weinstein, "Nonlinear Schrödinger equations and sharp interpolation estimates," *Commun. Math. Phys.*, vol. 87, pp. 567–576, 1983, doi: 10.1007/BF01208265.

[27] K. Yang, A. M. Swope, A. Gu, R. Chalamala, P. Song, S. Yu, S. Godil, R. J. Prenger, and A. Anandkumar, "LeanDojo: Theorem proving with retrieval-augmented language models," in *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[28] Y. Zhou, "Regularity criteria for the generalized viscous MHD equations," *Ann. Inst. H. Poincaré C Anal. Non Linéaire*, vol. 24, no. 3, pp. 491–505, 2007, doi: 10.1016/j.anihpc.2006.03.014.