Supplementary Information for “Learned variable-support priors accelerate symbolic regression of physical laws”

Chenxi He*[1] and Jingqiu Chen[2]

[1]Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, United Kingdom
[2]College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China

This Supplementary Information contains: (i) a selector-summary table that compares DenoisedSR with five classical feature selectors across AI-Feynman, SRSD-Feynman and noise levels (Supplementary Table S1); (ii) a threshold-sensitivity table demonstrating that the deployed $\tau = 0.10$ operating point is not cherry-picked (Supplementary Table S2); (iii) an internal-validation table reporting two additional, internally-curated suites under a different evaluation protocol (Supplementary Table S3); (iv) a backend integration summary (Supplementary Table S6); (v) the master backend $\times$ benchmark matrix (Supplementary Table S7); (vi) an 80-equation Feynman scale-up (Supplementary Table S8); (vii) controlled variable-name ablations (Supplementary Tables S4 and S5); and (viii) the SRSD-Feynman per-task precision diagnostic (Supplementary Fig. S1). All numbers reproduce from the released evaluation scripts at `https://github.com/ChenxiHeCam/DenoisedSR` (archived at `10.5281/zenodo.20584889`).

Supplementary Table S1. Summary of variable-support recovery across selectors and suites. Variable-support recall, with the fraction of tasks on which the selector attains *perfect* recall (every true variable retained). All selectors except DenoisedSR are run at an oracle top-k budget that hands them the true variable count for free. Recall on AI-Feynman is also reported under relative Gaussian label noise η; entries marked — were not part of the noise sweep.

	AI-Feynman (118)		SRSD (120)	AI-Feynman recall under noise η				
Selector	recall	perfect %	recall	0.01	0.05	0.10	0.20	0.30
Pearson $\|\rho\|$	0.768	45.8	—	0.772	0.769	0.771	0.763	0.768
Spearman $\|\rho\|$	0.776	45.8	—	0.773	0.770	0.770	0.757	0.754
mutual information	0.634	20.3	—	0.633	0.636	0.644	0.608	0.605
random-forest importance	0.812	44.9	—	0.815	0.822	0.815	0.807	0.799
Lasso (CV) *oracle k*	0.905	71.2	0.723	0.904	0.904	0.886	0.881	0.868
DenoisedSR (ours, τ=0.10)	**1.000**	**100.0**	**0.967**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

References

*Corresponding author: `ch2067@cam.ac.uk`

Supplementary Table S2. Threshold sensitivity of the deployed RF+GAT ensemble. Sweep of τ on AI-Feynman (118 laws, $q = 100$, 20 distractors). Recall remains 1.000 with 100% perfect-recall over $\tau \in [0.05, 0.30]$ (a sixfold range); only at $\tau \geq 0.50$ does recall start to decline. The deployed $\tau = 0.10$ is not cherry-picked.

τ	Recall	Precision	Perfect-recall (%)	Mean kept columns
0.05	1.000	0.705	100.0	9.03
0.08	1.000	0.740	100.0	8.81
0.10 (deployed)	**1.000**	**0.749**	**100.0**	**8.76**
0.15	1.000	0.766	100.0	8.67
0.20	1.000	0.769	100.0	8.66
0.30	1.000	0.780	100.0	8.60
0.50	0.998	0.873	99.2	6.31
0.75	0.981	0.991	93.2	3.87
1.00	0.959	1.000	86.4	3.72

Supplementary Table S3. Internal validation under a different protocol (not directly comparable to the four external benchmarks of Fig. ??). Evaluated on internally-curated physical-law records drawn from the same knowledge graph as the training pool, using each record's pre-sampled observation values rather than freshly sampled tables, and without random distractor padding. We report it as a separate scale check rather than a fifth external suite.

Set	Equations	Recall	Perfect-recall (%)
Internal physics-law set (v5_physics65)	1085	0.999	99.7
Internal OOD-91 curated subset	91	0.996	98.9

Supplementary Table S4. Variable filter is insensitive to physical names. Same data, same model, same threshold $\tau = 0.10$; only the variable names are changed between "named" (physical: m, v, q, ϵ, $\ldots$) and "anonymized" (x_0, x_1, $\ldots$). Across 12 scenarios on AI-Feynman 118, the physical-name prior contributes at most 0.3 percentage points of variable filter rate. The deployed RF+GAT ensemble is dominated by the random-forest column-statistics head; the graph-attention head's co-occurrence-graph contribution is largely vestigial for the variable task.

	Named (physical)		Anonymized (x_i)		Δ (named−anon)	
Scenario	recall	filter%	recall	filter%	recall	filter (pp)
baseline q=100, d=20	1.000	98.4	0.998	98.3	+0.002	+0.04
small q=50	1.000	97.0	0.998	97.0	+0.002	+0.04
very small q=20	1.000	90.8	0.998	90.6	+0.002	+0.18
noise $\eta = 0.30$	1.000	98.7	0.998	98.6	+0.002	+0.08
noise $\eta = 0.50$	1.000	98.6	0.998	98.5	+0.002	+0.10
noise $\eta = 1.00$	1.000	98.6	0.998	98.6	+0.002	+0.07
many distractors d=50	1.000	98.5	0.998	98.5	+0.002	+0.04
many distractors d=100	1.000	98.6	0.998	98.5	+0.002	+0.09
hard q=20, $\eta = 0.50$	1.000	90.9	0.998	90.7	+0.002	+0.27

Supplementary Table S5. Operator filter *is* sensitive to physical names. Same data, same operator predictor, varying threshold τ_{op}. Operator filter rate is the fraction of the 13 non-trivial operators (excluding the four always-on arithmetic operators) dropped from the search library. Physical names contribute +7 to +17 percentage points of operator recall—a much larger effect than for variables, because the operator-vs-data signal in $(\mathbf{x}, y)$ is weak and the physics-knowledge-graph prior is the main source of information.

τ_{op}	Named (physical)			Anonymized (x_i)			Δ recall
	recall	prec.	filter%	recall	prec.	filter%	(named−anon)
0.03	0.912	0.383	58.9	0.744	0.346	59.8	+0.168
0.10	0.825	0.465	67.9	0.726	0.380	65.6	+0.099
0.20	0.761	0.503	72.9	0.703	0.427	70.1	+0.058
0.30 (deployed)	**0.738**	**0.536**	**76.5**	**0.667**	**0.437**	**72.9**	+0.071
0.50	0.597	0.547	82.7	0.521	0.420	82.2	+0.076
0.70	0.407	0.532	88.9	0.285	0.376	90.4	+0.122

Supplementary Table S6. Backend integration summary. Mean exact-law recovery and mean held-out R^2 on the 30-equation Feynman headline subset ($n_{\text{dist}} = 20$, $q = 100$). Errors are ±s.d. over three random seeds. The variable prior shrinks the candidate-column set from 23.2 to 3.4 in both backends.

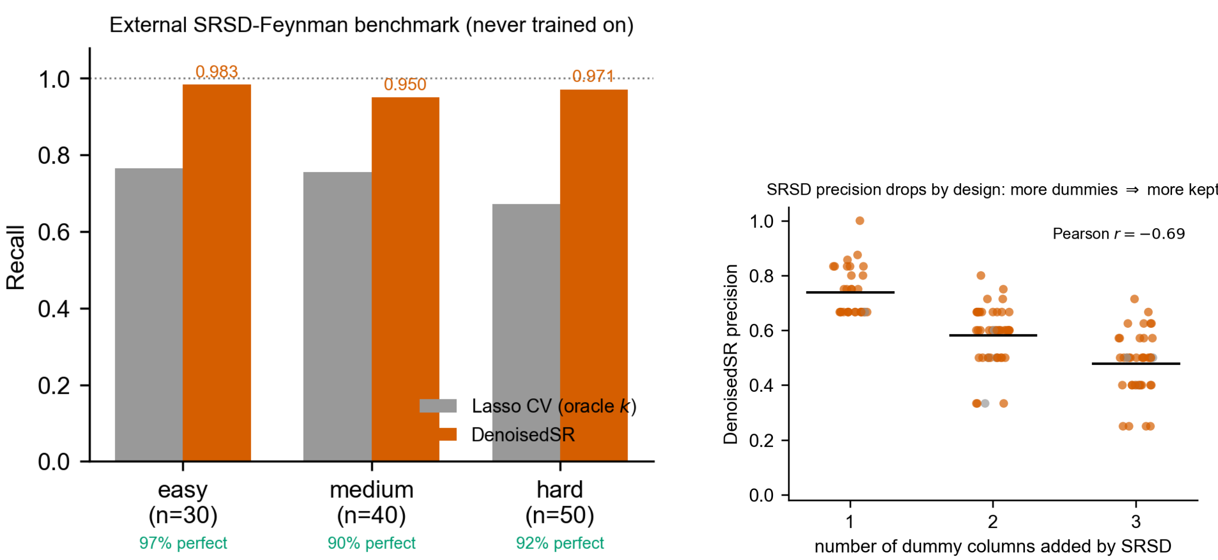
Condition	Exact recovery (%)	Mean held-out R^2
PySR (10 s budget, full 17-op library)		
full search	42.2 ± 13.5	0.906 ± 0.045
+ DenoisedSR variable prior	60.0 ± 0.0	0.980 ± 0.011
+ DenoisedSR variable + operator prior	60.0 ± 11.5	0.979 ± 0.011
gplearn (20 generations, population 500, 10-op library)		
full search	3.3 ± 3.3	0.475 ± 0.083
+ DenoisedSR variable prior	17.8 ± 1.9	0.763 ± 0.029
PSE (parallel GPU enumeration, 30 s budget, 9-op library, single seed)		
full search	0	0.184
+ DenoisedSR variable prior	40	0.895

Supplementary Table S7. Master backend × benchmark matrix. Exact-law recovery (% with held-out $R^2 \geq 0.999$) on five physical-symbolic benchmarks for three structurally different SR backends, with and without the DenoisedSR variable prior. Cells show *full backend → +DenoisedSR*; gap in percentage points. Multi-seed entries are mean±s.d.; single-seed entries omit the standard deviation. PySR-Feynman118 is over three random seeds; gplearn entries on Feynman are over three seeds; remaining cells are single seed. PSE-SRSD is not evaluated because PSE's float32 tensor representation underflows on SRSD's $\sim 10^{60}$-magnitude physical-unit targets (a backend-data interaction, unrelated to the prior). Bold marks the larger of the two values in each row.

Benchmark	Distractors	PySR *full → +prior* (Δ)	gplearn *full → +prior* (Δ)	PSE *full → +prior* (Δ)
AI-Feynman 118 (3 seeds)	20 random	46.3±3.0 → **57.3±1.8** (+11)	0.8 → **11.0** (+10)	8.5 → **42.4** (+34)
Strogatz 15	20 random	60.0 → **66.7** (+7)	0.0 → 0.0 (0)	6.7 → **53.3** (+47)
Nguyen 12	20 random	**91.7** → **91.7** (0)	8.3 → 8.3 (0)	33.3 → **100.0** (+67)
SRSD-Feynman 120	∼2 plausible	**54.2** → 50.0 (–4)	30.0 → 30.0 (0)	—
SRSD-Feynman 120	+ 20 random	42.5 → **54.2** (+12)	25.0 → **29.2** (+4)	—

Supplementary Table S8. Headline result scales to 80 Feynman equations. Repeating the PySR comparison at the same configuration ($n_{\text{dist}} = 20$, $q = 100$, $\tau = 0.10$, single seed) on the first 80 Feynman tasks (chosen by $2 \leq |\text{features}| \leq 9$) rather than the headline 30. The gain is preserved at larger scale; full PySR's higher absolute rate at this scale (54% vs. 30–57% across seeds at n=30) is consistent with the 30-equation subset including a higher fraction of harder formulas.

Condition	Exact (%)	Mean held-out R^2	Variable recall
full PySR (all 23 cols, 17 ops)	54	0.900	—
+ DenoisedSR variable prior (3.4 cols)	66	0.981	1.000
+ DenoisedSR variable + operator prior	68	0.981	1.000



Supplementary Fig. S1. Detail on SRSD-Feynman. *Left*: recall on each SRSD difficulty split (30 easy, 40 medium, 50 hard) for DenoisedSR vs. Lasso at oracle top-k; DenoisedSR exceeds 0.95 recall on every split. *Right*: diagnostic of the apparent precision drop on SRSD (0.59 overall). Each marker is one of the 120 SRSD tasks (orange = perfect recall, grey = imperfect recall); horizontal bars show the binned mean precision. Precision falls monotonically with the number of dummy columns SRSD adds to the task (Pearson $r = -0.69$) because the front-end at τ=0.10 retains physically plausible dummies rather than discarding them; on tasks with no dummies our precision is near 1.