NEURAL INFORMATION PROCESSING SYSTEMS

Evolution Fine-Tuning: Learning to Discover Across 371 Optimization Tasks

NeurIPS 2026

Young-Jun Lee[1], Seungone Kim[2], Minki Kang[3], Alistair Cheong Liang Chuen[2], Zerui Chen[4], Seungho Han[5], Taehee Jung[6], Dongyeop Kang[1]

[1]University of Minnesota, [2]Carnegie Mellon University, [3]KAIST, [4]University of Cambridge, [5]Hanyang University, [6]Amazon

UNIVERSITY OF MINNESOTA

Carnegie Mellon University
Language Technologies Institute

KAIST

UNIVERSITY OF CAMBRIDGE

HANYANG UNIVERSITY 한양 1939

amazon

Would experience designing faster GPU kernels also help close in on a longstanding open mathematical conjecture?

Would experience designing faster GPU kernels also help close in on a longstanding open mathematical conjecture?

= Cross-Discovery Transfer

Would experience designing faster GPU kernels also help close in on a longstanding open mathematical conjecture?

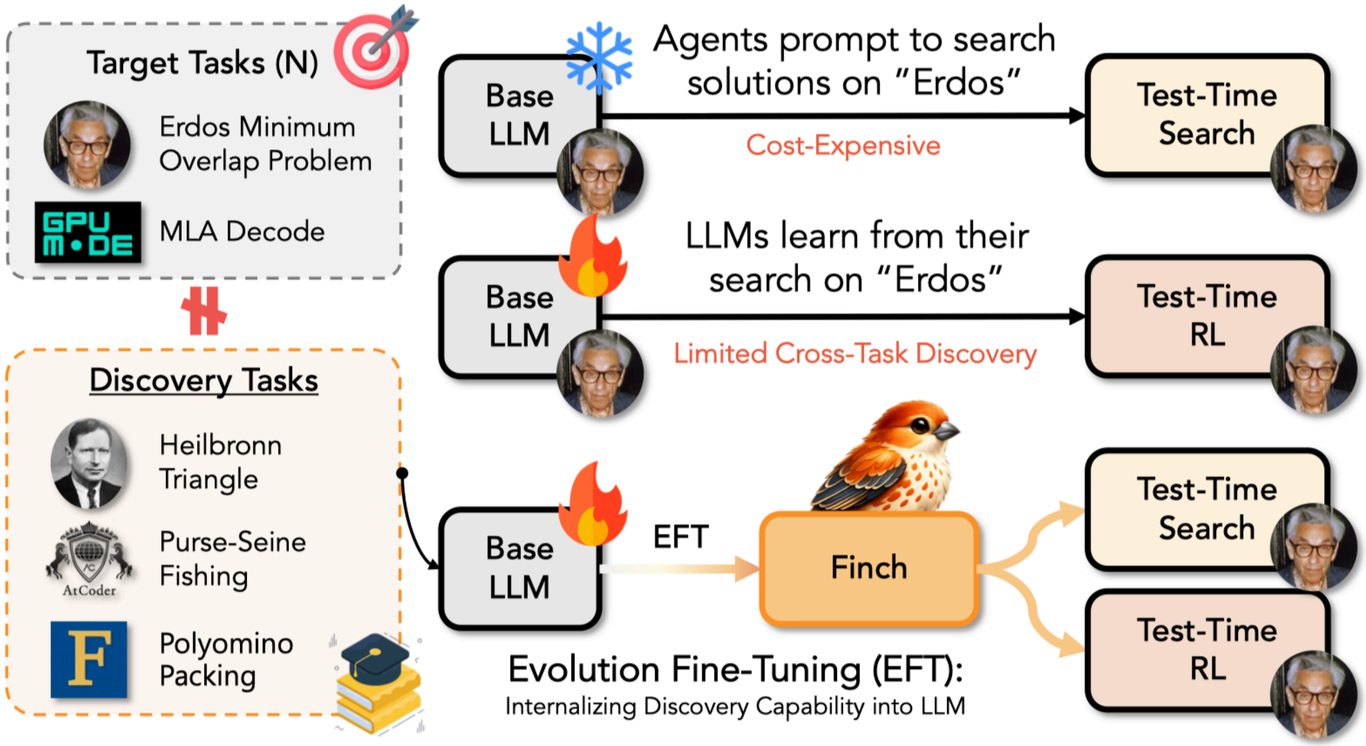
= Cross-Discovery Transfer

💊 **Internalizing Discovery Capability into LLM**

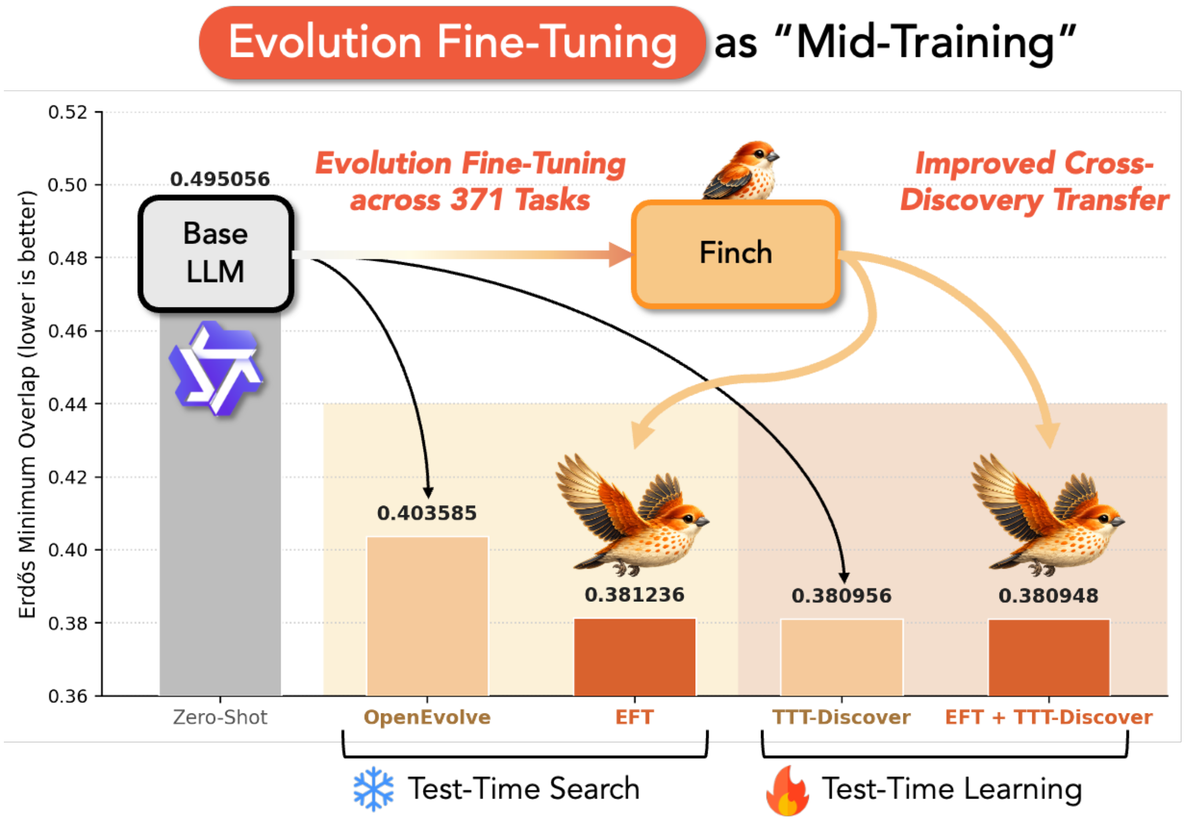
Evolution Fine-Tuning:



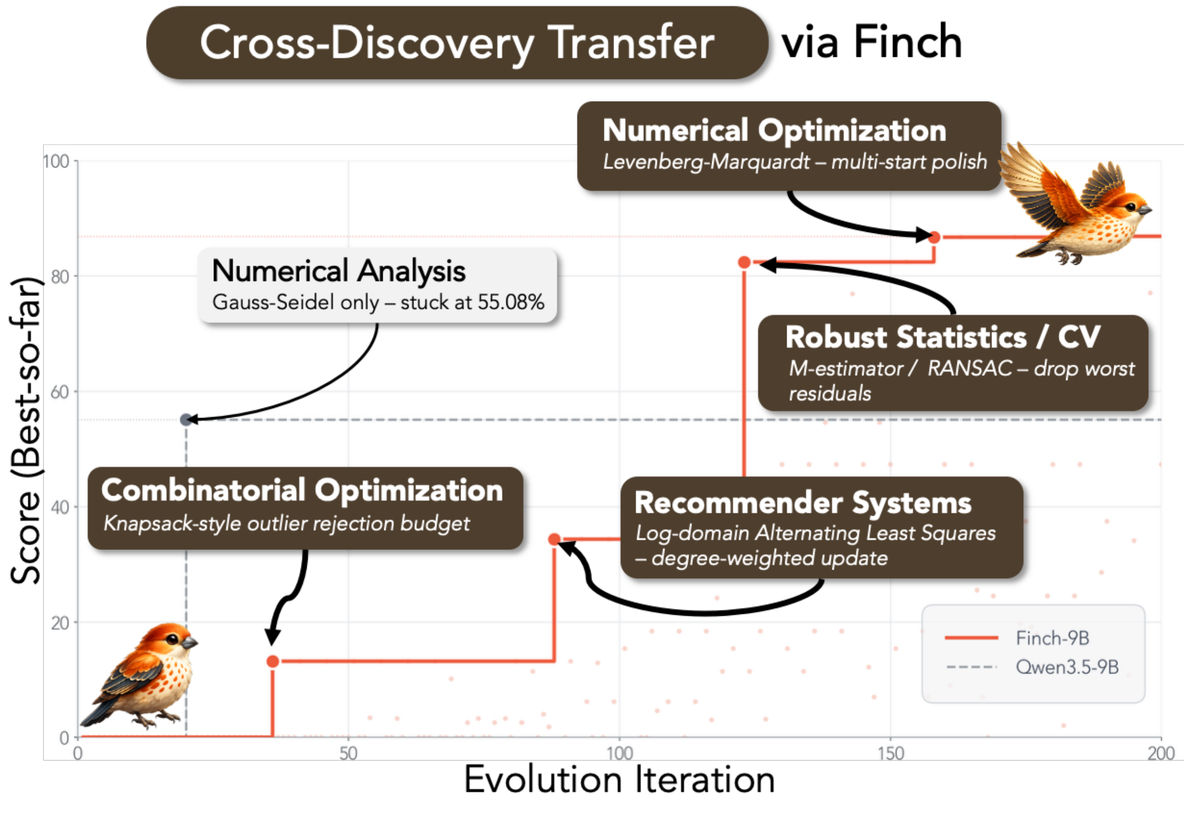
A ***mid-training*** paradigm that treats the ***trajectories*** of search runs as the supervision signal



What Are the Key Advantages?



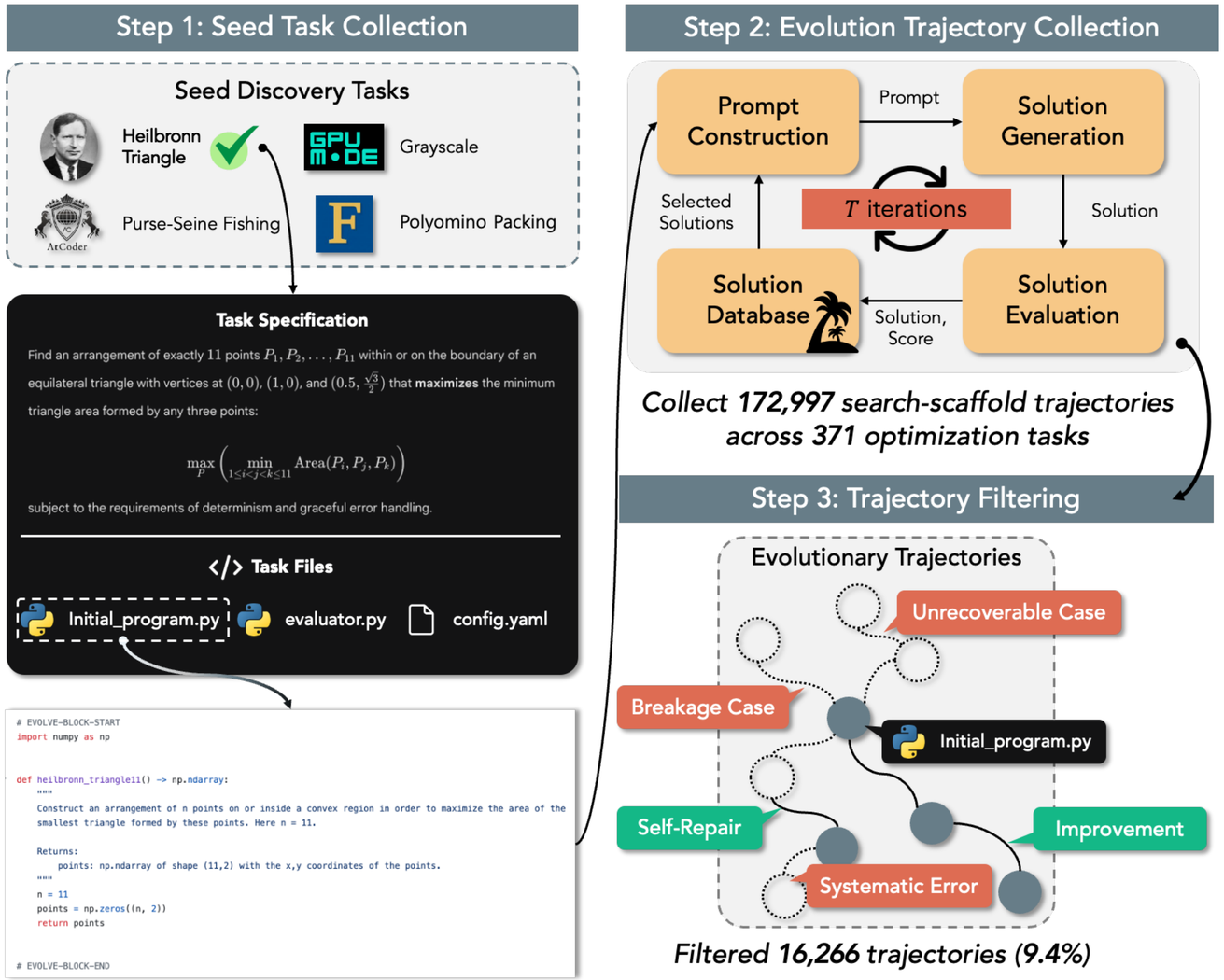
Erdos Minimum Overlap Problem



NP-hard competitive programming (CALICO, UC Berkeley contest)

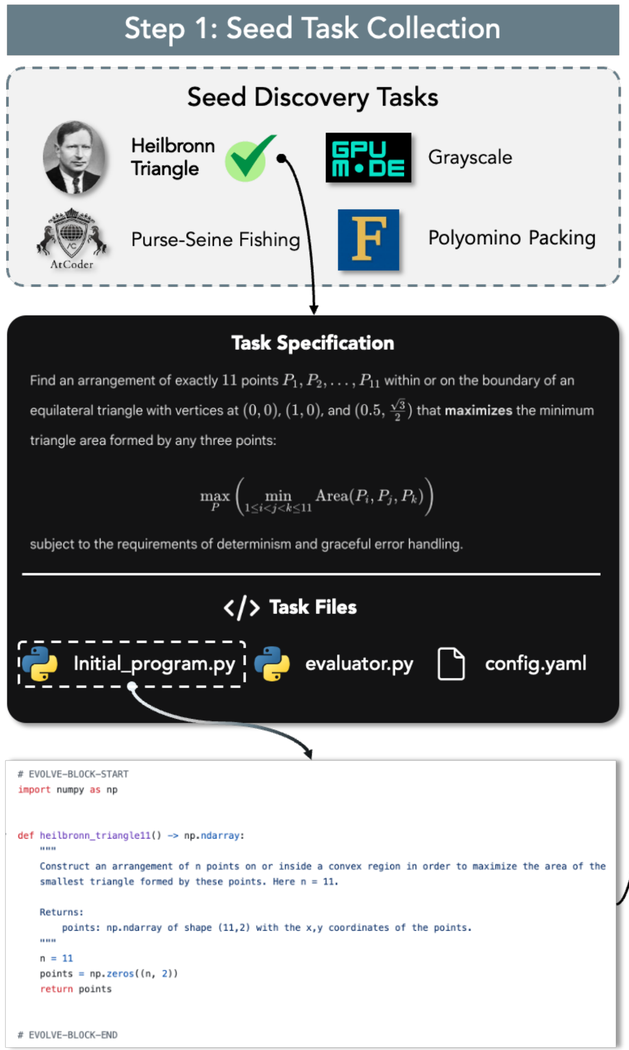
Finch Collection:

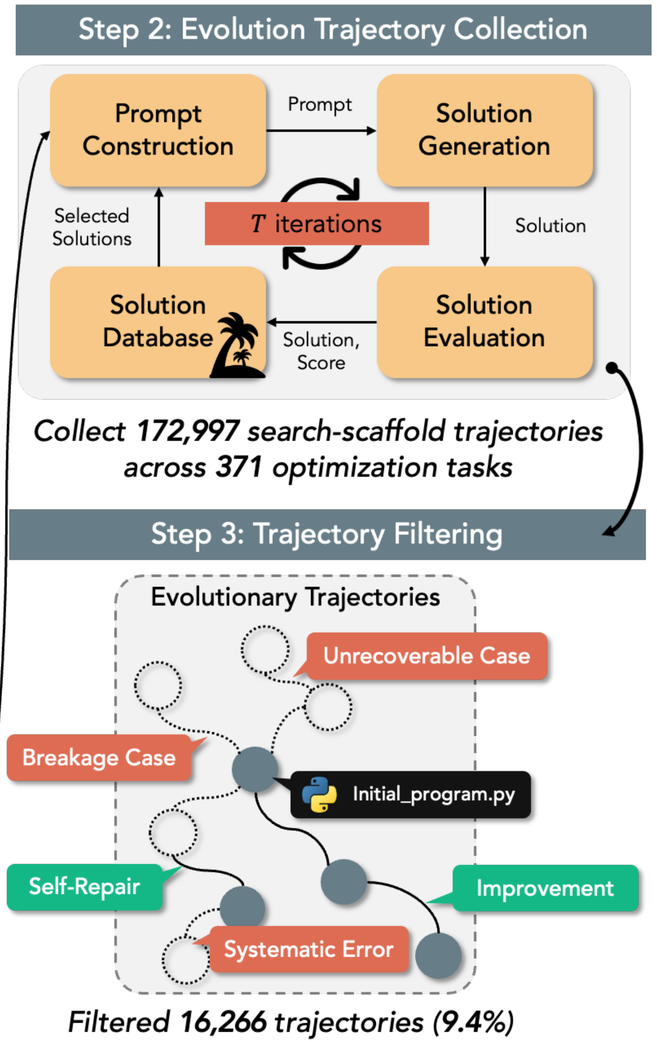
156K-Scale Trajectories across 371 Optimization Tasks

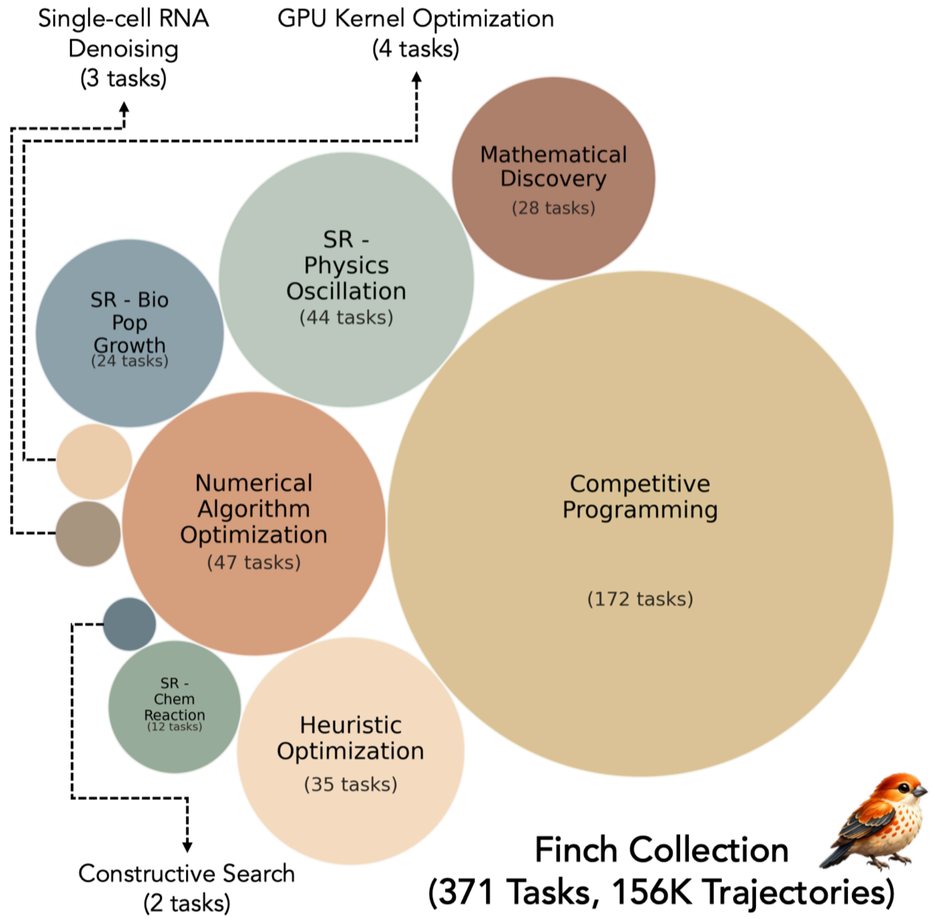


Finch Collection:

156K-Scale Trajectories across 371 Optimization Tasks







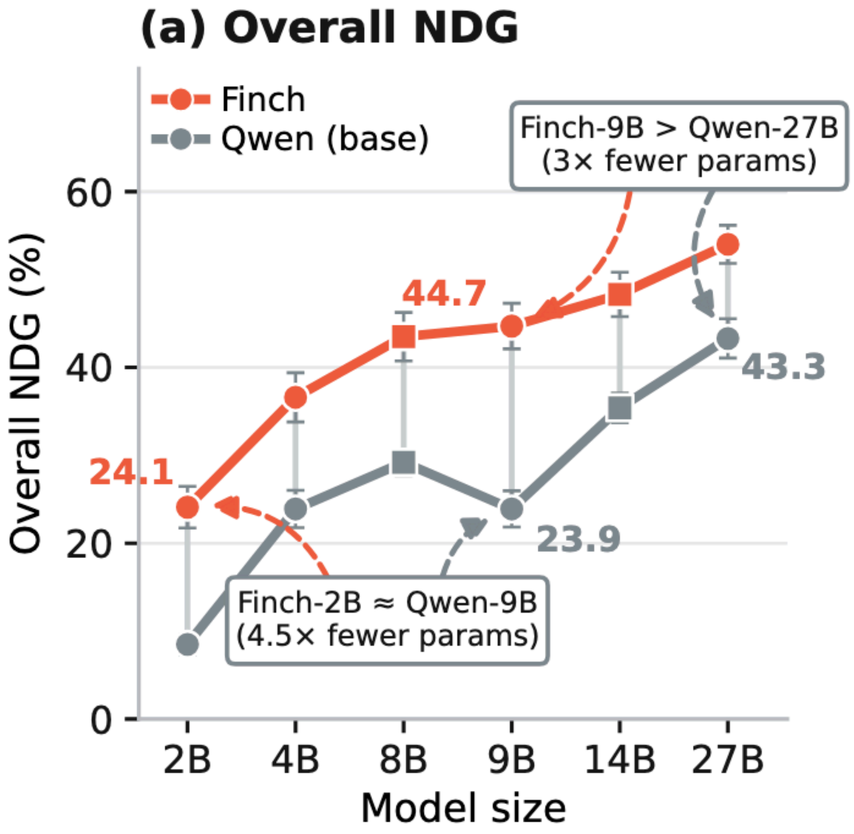
Finch: 

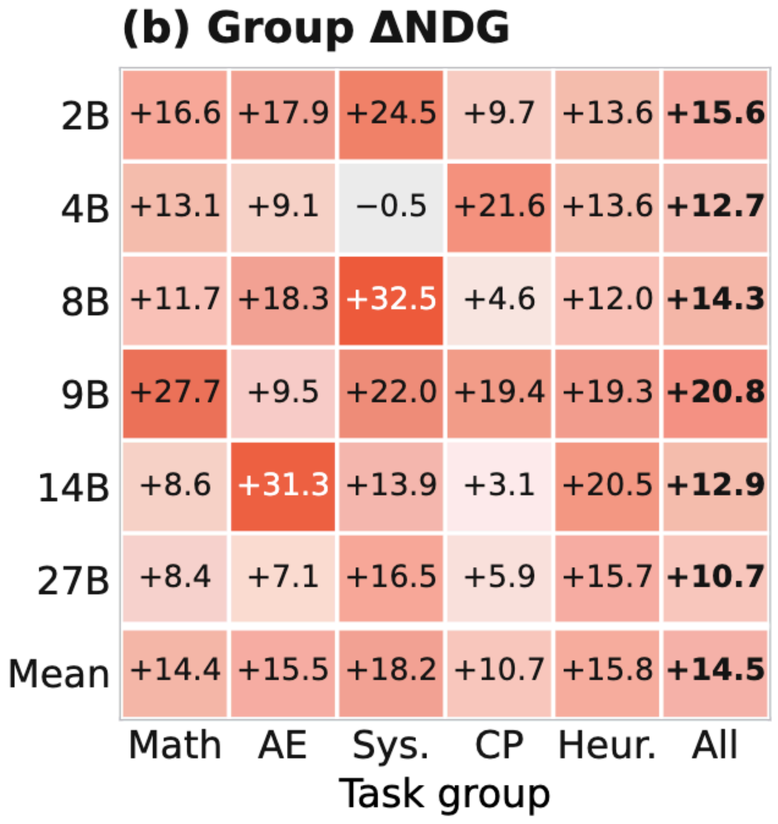
Evolution Fine-Tuned Language Model

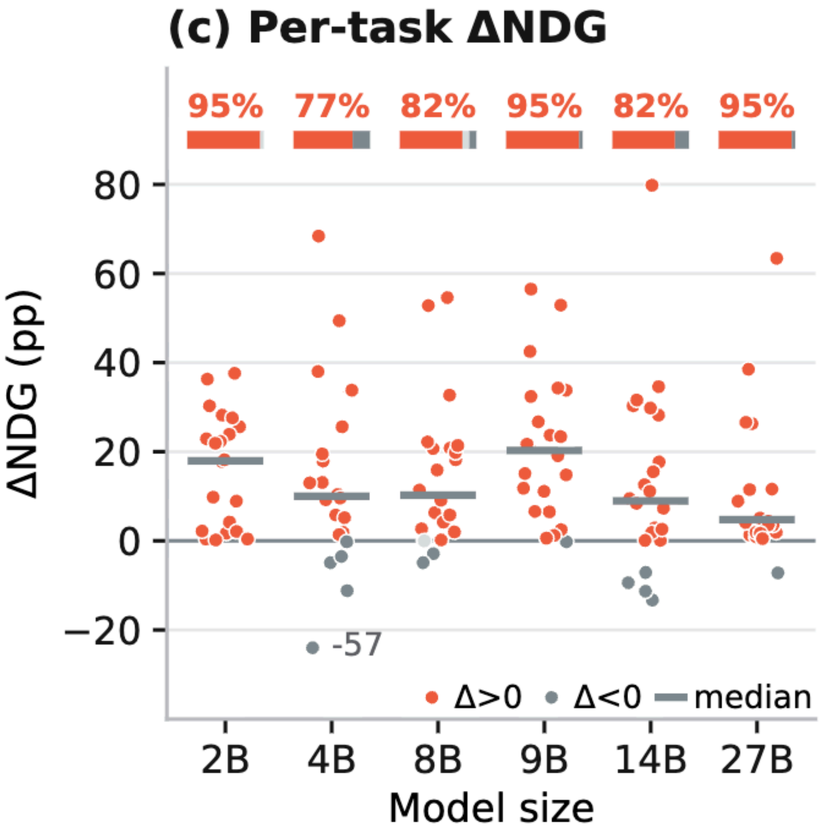
- **Base Models**: Qwen3.5-{2, 4, 9, 27}B & Qwen3-{8, 14}B
- **355** Training Tasks / **16** Held-out Tasks
- For **EFT**, we train base models on "**Improved**" trajectories across 355 tasks using **cross-entropy loss**
- For **offline RL**, we train Finch on both "**Improved**" & "**Regressed**" trajectories across 355 tasks using **KTO**
- For training, we employ the **LLaMA-Factory** framework

Results [1/3]

EFT improves cross-task discovery generalization across all model scales







Results [2/3]

EFT Complements Offline RL and Test-Time RL

Table 3: Effect of Offline RL (CP: avg. competitive programming score).

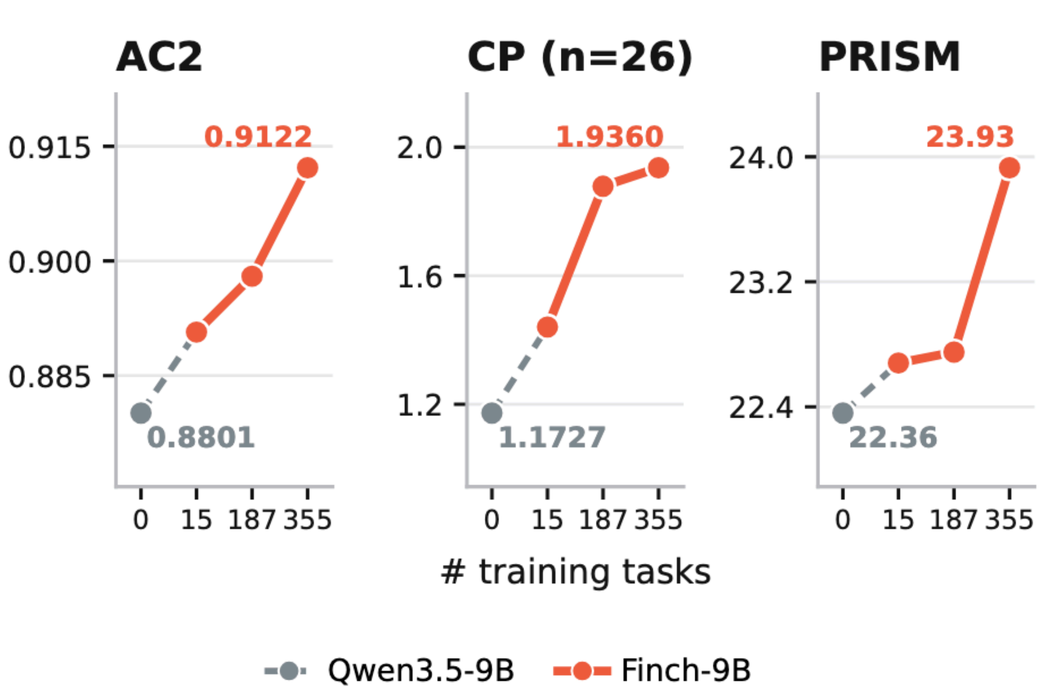
Model	Erdős (↓)	AC1 (↓)	AC2 (↑)	CP (↑)
Best Human	**0.380927**	1.5097	0.9015	–
Qwen3.5-4B	0.416924	1.5186	0.8802	14.52
$\mathcal{F}$inch-4B	0.386460	1.5173	0.8933	31.97
$\mathcal{F}$inch-4B + KTO	0.381809	1.5151	0.9121	**36.30**
Qwen3-8B	0.403585	1.5177	0.8980	24.21
$\mathcal{F}$inch-8B	**0.381236**	1.5154	0.9001	24.56
$\mathcal{F}$inch-8B + KTO	0.381596	**1.5089**	**0.9146**	**37.30**

Table 4: Performance of $\mathcal{F}$inch combined with online (test-time) RL scaffolds across math optimization tasks.

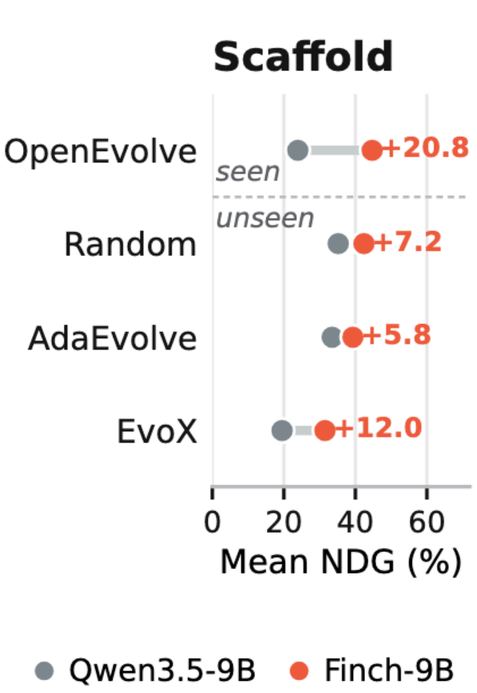
Scaffold	Model	Erdős (↓)	CP (n=26) (↑)	CP (n=32) (↑)
ThetaEvolve	R1-Qwen3-8B	-	**2.635983**	-
TTT-Discover	GPT-OSS-120B	0.380876	-	-
	Qwen3-8B	**0.380932**	**2.635983**	2.939572
nanodiscover	Qwen3-8B	0.380956	**2.635983**	**2.939573**
	$\mathcal{F}$inch-8B	**0.380948**	**2.635983**	**2.939573**

Results [3/3]

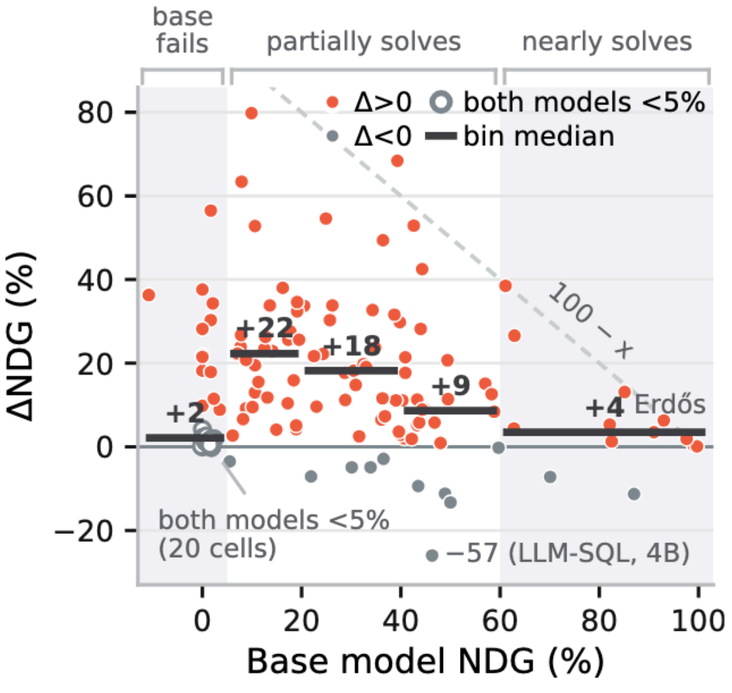
EFT Scales and Generalizes Across Scaffolds



(a) Scaling with the number of training tasks.



(b) Unseen scaffolds.



(c) Gain vs. base-model NDG.

Thank you

https://passing2961.github.io/

Please come visit our poster session!