From *f(x)* and *g(x)* to *f(g(x))*:
LLMs Learn New Skills in RL by Composing Old Ones

Lifan Yuan*, *Weize Chen*, Yuchen Zhang, Ganqu Cui, Hanbin Wang, Ziming You, Ning Ding, Zhiyuan Liu, Maosong Sun, Hao Peng*

Oct 14 @ ASAP Seminar

Overview

- Why do we care about if LLMs learn skills in RL
- How to answer this question in a clean way
- Findings

Ongoing Debate: Does RL Teach New Skills?

- Traditional view: RL teach LLMs genuinely new skills

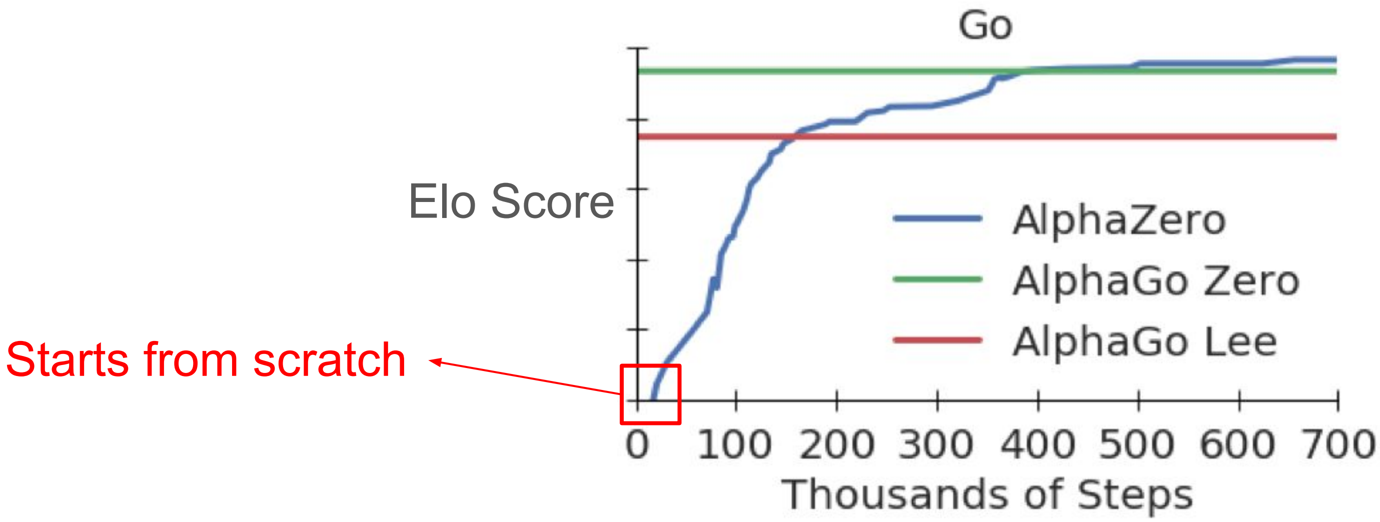
- Recent findings: RL mainly activates existing patterns [1,2]

[1] Zhao et al. 2025. Echo Chamber: RL Post-training Amplifies Behaviors Learned in Pretraining.

[2] Yue et al. 2025. Does Reinforcement Learning Really Incentivize Reasoning Capacity in LLMs Beyond the Base Model?

Ongoing Debate: Does RL Teach New Skills?

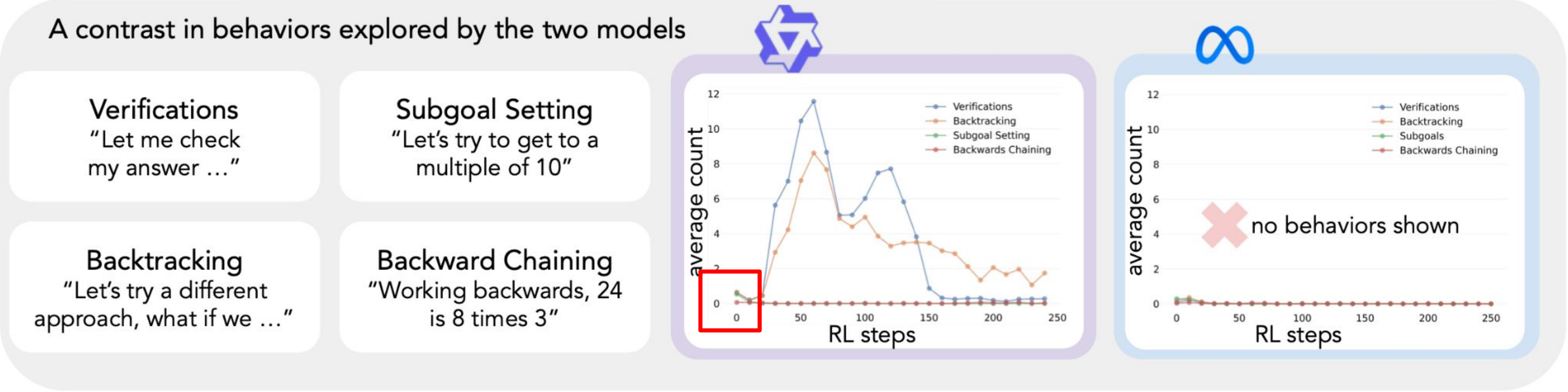
- Models can be trained from scratch and ended up beating humans in Go.



Silver et al. 2017. Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm

Ongoing Debate: Does RL Teach New Skills?

- Models can be trained from scratch and ended up beating humans in Go.
- Work in LLMs suggests there is no “Aha moment” but pattern activation



Gandhi et al. 2025. Cognitive Behaviors that Enable Self-Improving Reasoners, or, Four Habits of Highly Effective STaRs

Ongoing Debate: Does RL Teach New Skills?

- Models can be trained from scratch and ended up beating humans in Go.
- Work in LLMs suggests there is no “Aha moment” but pattern activation

Difference?

Vast action space $\longrightarrow$ Pretrained model prior

- Reasonable trajectories
- Constrained search space

Why is the problem important?

It impacts important decisions in the life cycle of LLMs:

- If RL activates:
  better pre-training for more capable base models

- If RL teaches:
  refining the RL process itself to unlock novel skills

Research Questions

- Does RL teach new skills to LLMs?

- If so, how to incentivize it?

- Are the skills generalizable?

Research Questions

- Does RL teach new skills to LLMs?

  Yes, by the means of composing old skills
- If so, how to incentivize it?

  Include explicit incentive for composition
- Are the skills generalizable?

  Generalizes to held-out evaluation, more difficult problems, and even a different task

Intuition

Humans acquire cognitive skills by:

1. Learn basic knowledge
2. **Compose** sequences of actions into a single one
3. Internalize the new one

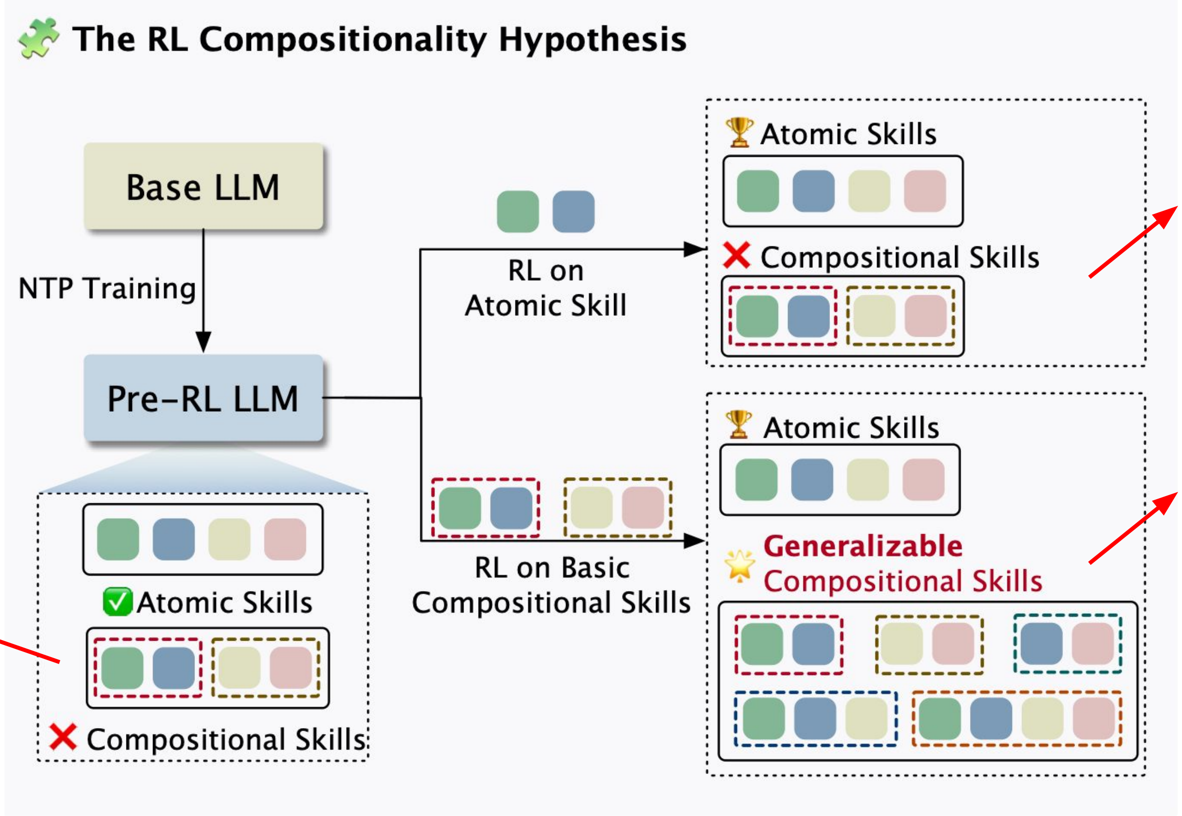
Anderson 1982. Acquisition of Cognitive Skill

RL Compositionality Hypothesis

Hypothesis:

Once a model has acquired the necessary atomic, non-decomposable skills for a task through NTP training, RL with proper incentivization can teach the model to learn new skills by composing atomic skills into more complex capabilities.

A Preview of Results



SFT only does not work

Only including atomic skill in RL does not work

Explicitly incentivizing composition in RL **works!**

Overview

- Why do we care about if LLMs learn skills in RL
- How to answer this question in a clean way
- Findings

Methodology: A Controlled Synthetic Framework

It is very important to reduce or eliminate spurious correlation between the base model and evaluation task, and to clearly distinguish atomic skills from compositional one.

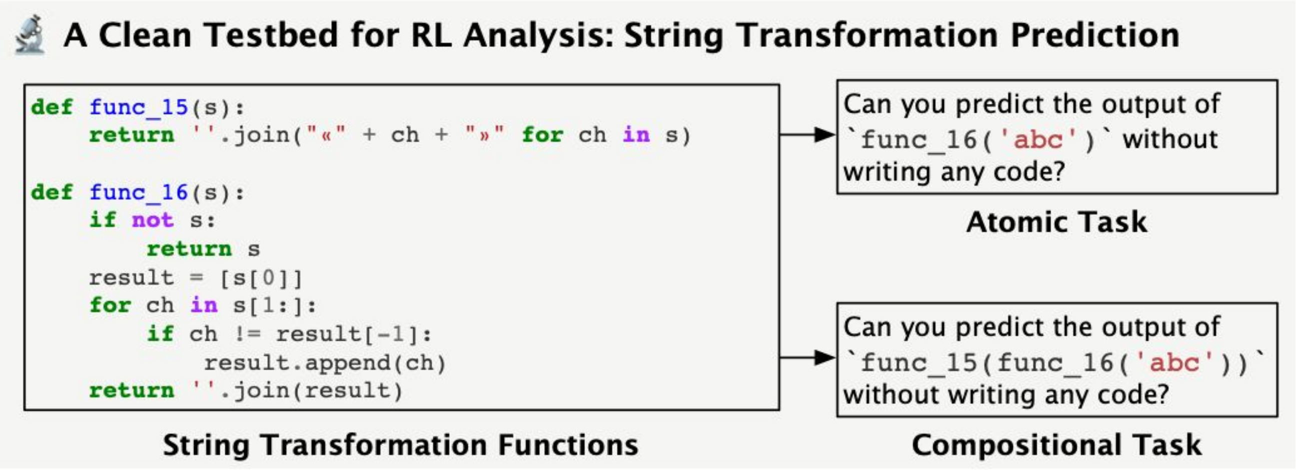
To this end, we need a framework where:

- The data does not appear in LLM’s pre-training
- Atomic skills are well-defined
- Task difficulty can be controlled

Methodology: A Controlled Synthetic Framework

Task Definition: String Transformation Prediction

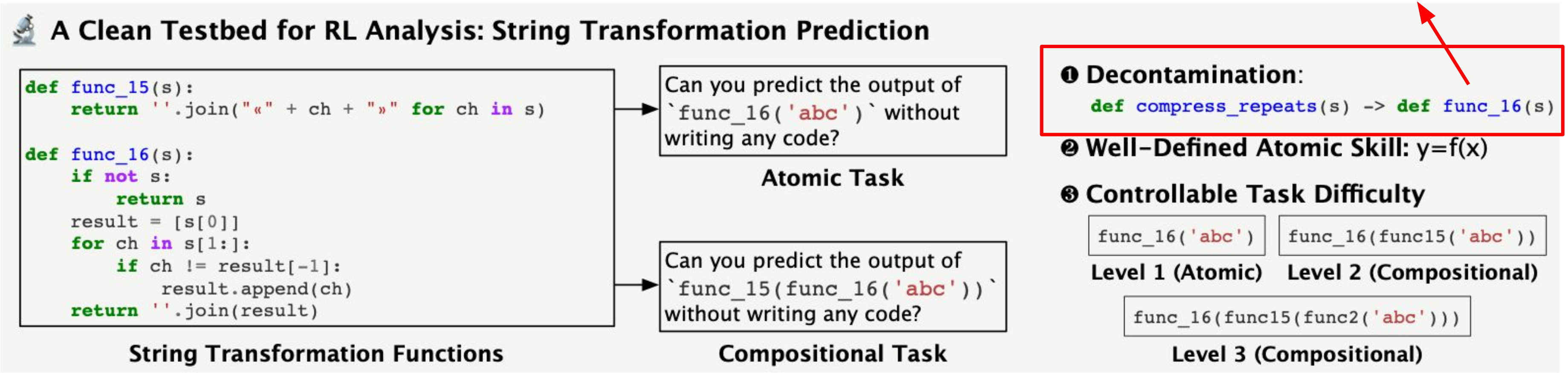
Given string input *x*, deterministic string transformations *f* and *g*, predicting expected output, e.g., $y=f(g(x))$.



Methodology: A Controlled Synthetic Framework

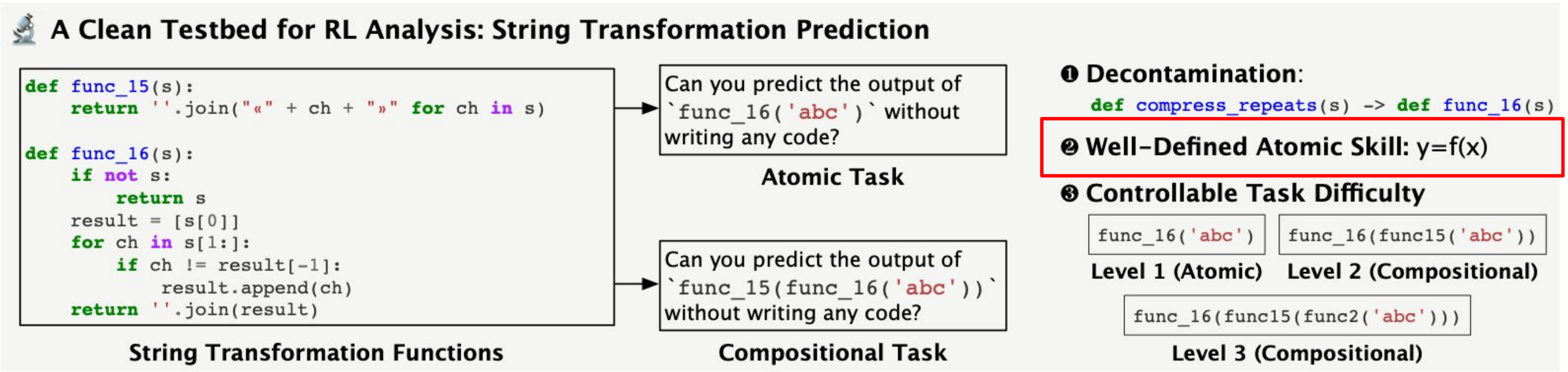
- Decontaminated Evaluation
- Well-defined Atomic and Compositional Skills
- Controllable Difficulty

Meaning less function name



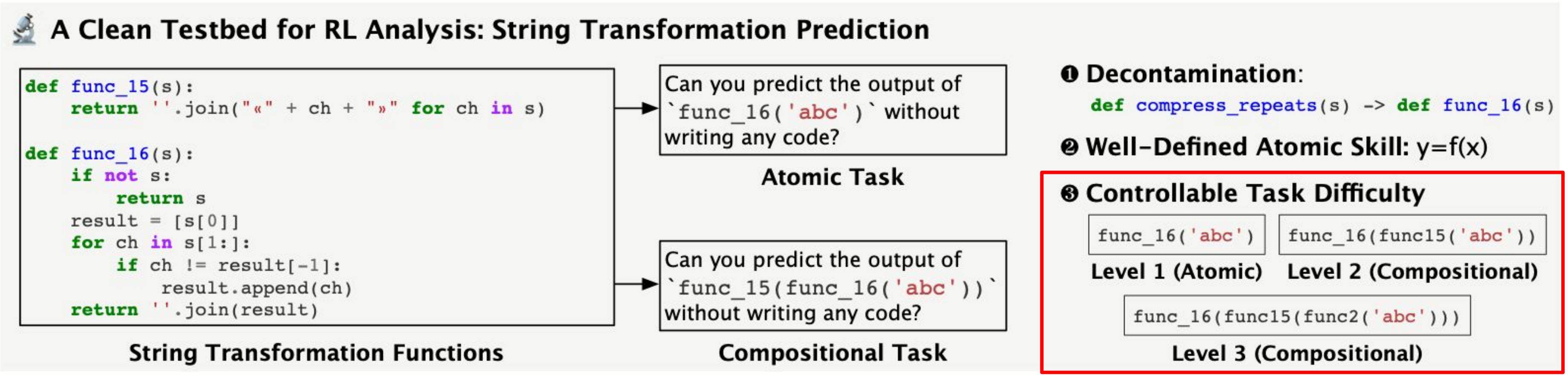
Methodology: A Controlled Synthetic Framework

- Decontaminated Evaluation
- Well-defined Atomic and Compositional Skills
- Controllable Difficulty



Methodology: A Controlled Synthetic Framework

- Decontaminated Evaluation
- Well-defined Atomic and Compositional Skills
- Controllable Difficulty



Two-Stage Training Setup Separating Atomic Skill Acquisition

Stage 1: Atomic Skill Training (Rejection Fine-Tuning, RFT)

Given full definition and string input, sample rollouts and train on correct ones.

Function Definition

```
You are given a code:

def func_16(s):
    """Remove adjacent duplicate characters (compress repeats)."""
    if not s:
        return s
    result = [s[0]]
    for ch in s[1:]:
        if ch != result[-1]:
            result.append(ch)
    return ''.join(result)

def main_solution(x):
    return func_16(x)

Can you predict the output of `main_solution("tihess")` without writing any code? Please reason and put
 your final answer in the following json format: {"output": <your output>}, where <your output> should
be the final string.
```

Input String

Two-Stage Training Setup Separating Atomic Skill Acquisition

Stage 2: Compositional Skill Training via Either RFT or RL

Predict transformation outputs **with function definition hidden**.

Example prompt for Stage 2 Level 1 training

Function definition is hidden

String input

```
You are given a code:

def main_solution(x):
    return func_16(x)

Can you predict the output of `main_solution("tiheass")` without writing any code? Please reason and
put your final answer in the following json format: {"output": <your output>}, where <your output>
should be the final string.
```

Example prompt for Stage 2 Level 2 training

Level 2

```
You are given a code:

def main_solution(x):
    return func_2(func_16(x), 3)

Can you predict the output of `main_solution("tiheass")` without writing any code? Please reason and
put your final answer in the following json format: {"output": <your output>}, where <your output>
should be the final string.
```

Evaluation Setup

Three types of generalization:

- **Held-out**

  We have 25 atomic skills in total, all are shown in Stage 1, but only 13 are shown in Stage 2.

  I.e., train on composition of func_1~func13, test on composition of func_14~func_25

- **Easy-to-Hard**

  Stage 1 on Level 1, Stage 2 on Level 1 or 2, Test on Level 1~6.

Overview

- Why do we care about if LLMs learn skills in RL
- How to answer this question in a clean way
- Findings

Finding 1: LLMs Acquire New Compositional Skills during RL

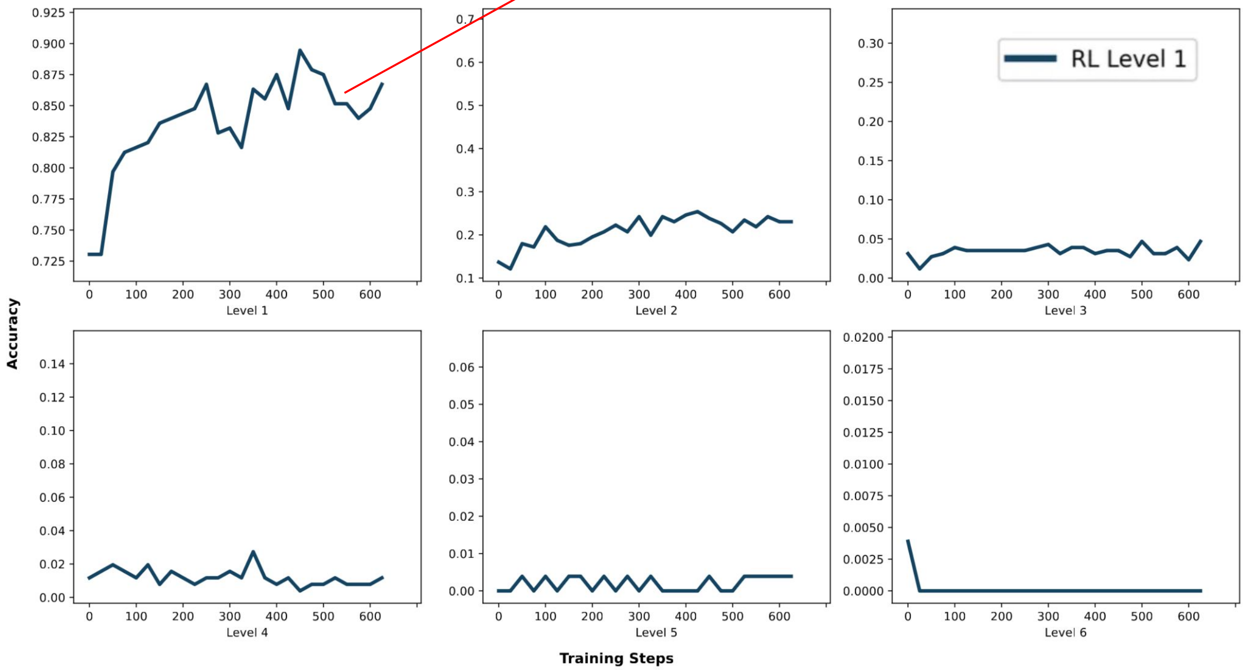
Different **data** setup in Stage 2:

1. RL on Level 1 problems only
2. RL on Level 2 problems only
3. RL on both Level 1 and Level 2 problems

Note that training on Level 1 problems do not explicitly incentivize composition

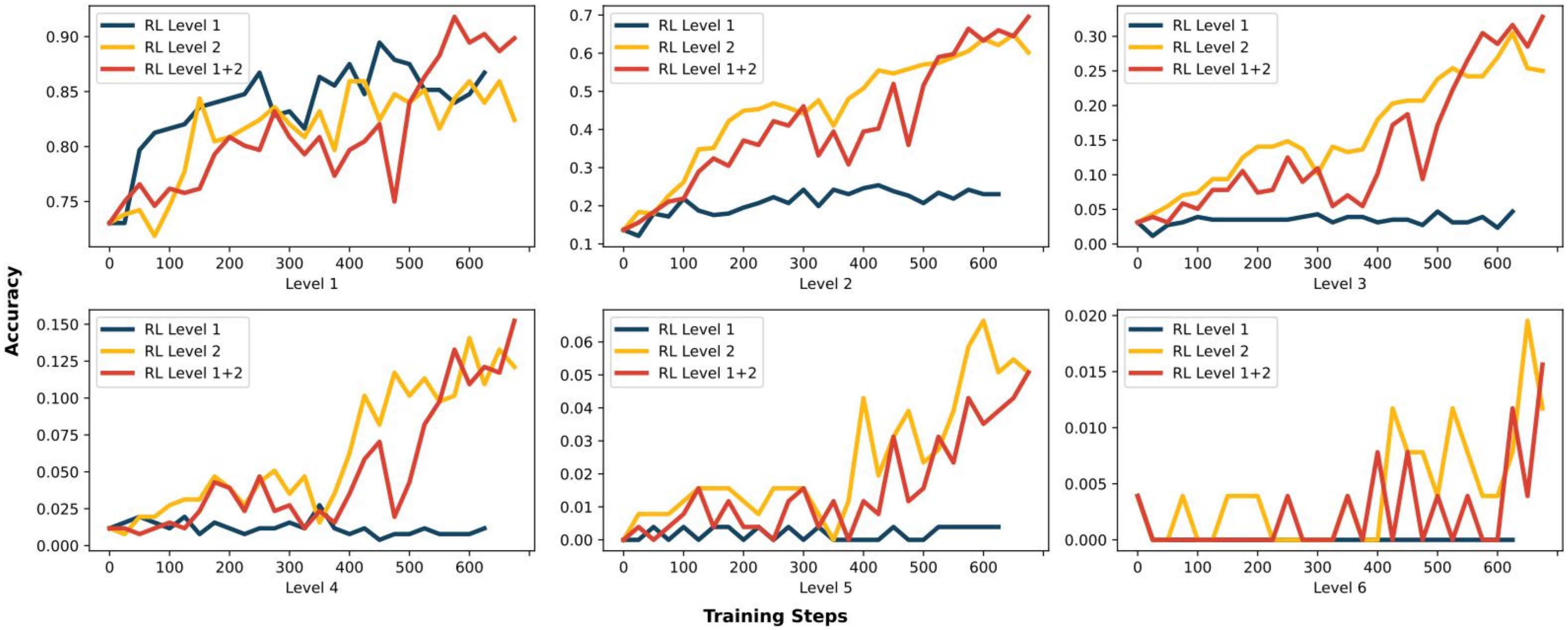
Finding 1: LLMs Acquire New Compositional Skills during RL

RL on atomic data leads to very high accuracy on atomic tasks (90%), but it is **insufficient** for learning effective composition.



Finding 1: LLMs Acquire New Compositional Skills during RL

RL on compositional data teaches new skills that generalize to unseen compositions of known atomic skills.



Finding 2: RL is the Key Ingredient to the New Compositional Skills

Previous experiments show that:

- RL on atomic data ❌
- RL on compositional data ✅

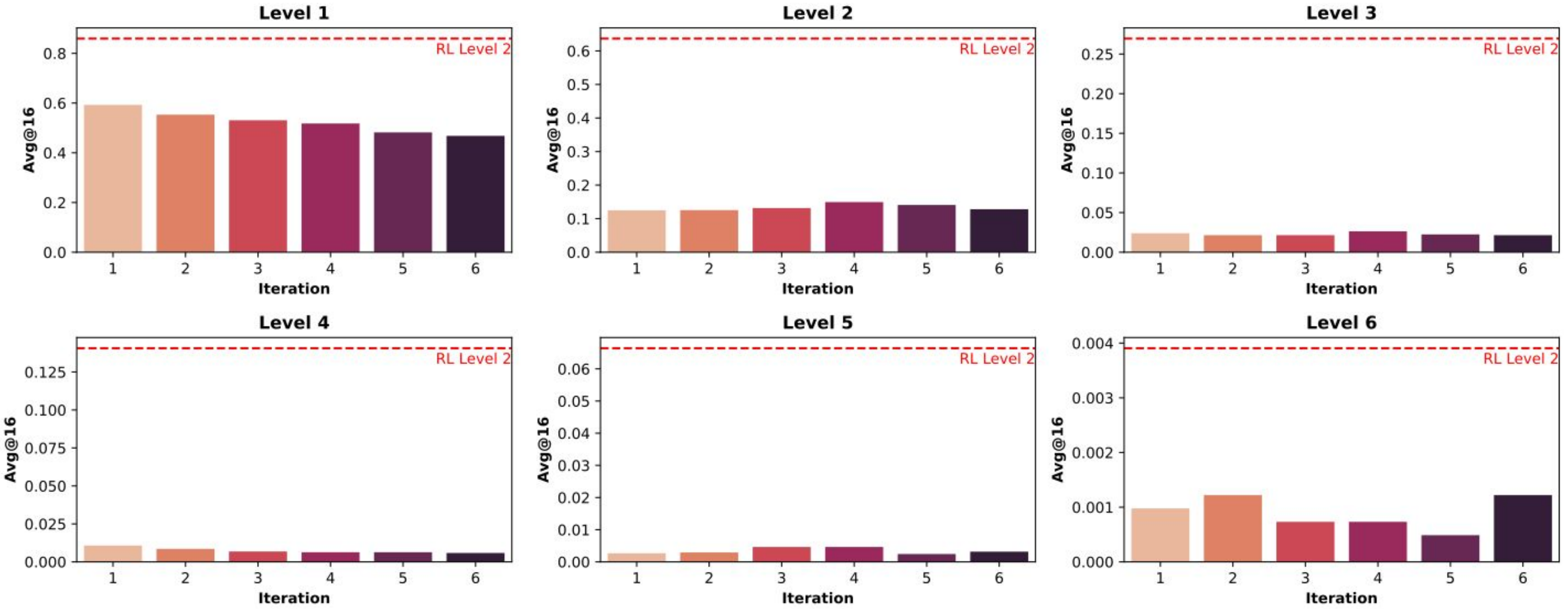
-> compositional data is important!

But is RL necessary?? Can a supervised method achieve the same results?

Compare RL with iterative RFT on Level 2 compositional data.

Finding 2: RL is the Key Ingredient to the New Compositional Skills

RFT, even with compositional data, is **suboptimal** for learning compositional skills, while RL is an important factor in learning generalizable compositional skills besides the data.



Finding 3: Compositional Skills Learned in RL Are Transferable, but Atomic Skills Are Prerequisites

Collecting compositional RL data for every new domain is impractical

> Test the transferability of the learned compositional skill.

We hypothesize that the compositional skills are transferable from Task A to Task B if:

1. The model has learned **atomic skills for Task B**.

2. The **compositional skill** has been learned through **RL on Task A**.

Use string transformation prediction as source task for RL, and countdown as target task:

User: Using the numbers [19, 36, 55, 7], create an equation that equals 65.

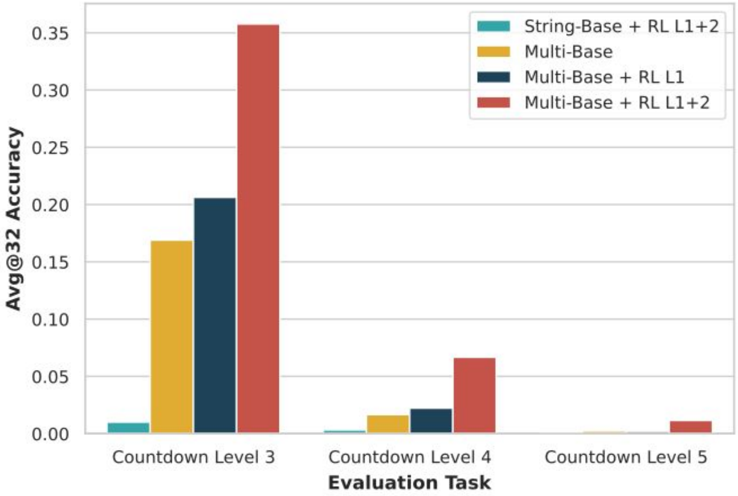
Level 4, atomic is Level 2

Finding 3: Compositional Skills Learned in RL Are Transferable, but Atomic Skills Are Prerequisites

		Stage 1		Stage 2	
	Model Configuration	**String Atomic RFT**	**Countdown Atomic RFT**	**String Atomic RL**	**String Comp. RL**
Abalate atomic skills in target task	String-Base + RL L1+2	✓	×	×	✓
Abalate Comp RL in source task	Multi-Base	✓	✓	×	×
	Multi-Base + RL L1	✓	✓	✓	×
Our ideal setup	Multi-Base + RL L1+2	✓	✓	×	✓

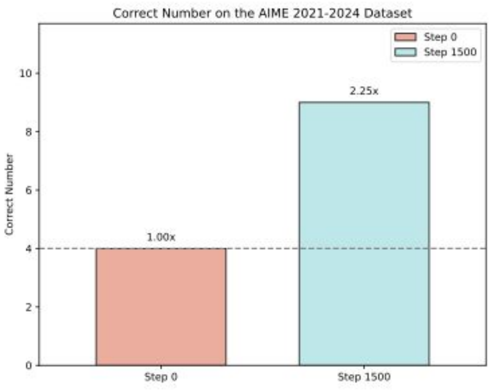
Multi-Base + RL L1+2 performs much better than the other three.

Compositional skills learned through RL are transferable to a different task where the model possesses the atomic skills.

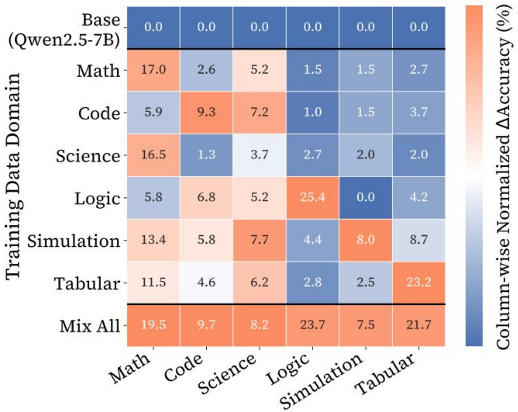


Finding 3: Compositional Skills Learned in RL Are Transferable, but Atomic Skills Are Prerequisites

Implications on existing work that demonstrates the generalizability of RL



Logic-RL [1] observes performance boost on unrelated math problems after training on logic puzzles



Guru models [2] show that domains with better exposure in pre-training data benefit more from cross-domain generalization

They may both be accounted for the fact that modern LLMs **have already gained necessary atomic skills** during the large-scale pretraining.

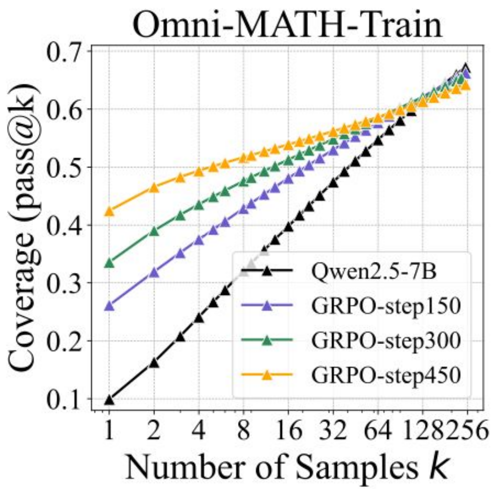
Therefore, incentivizing compositional skills by RL from another task helps combine task-specific skills in those tasks more effectively.

[1] Xie et al. 2025. Logic-RL: Unleashing LLM Reasoning with Rule-Based Reinforcement Learning
[2] Cheng et al. 2025. Revisiting Reinforcement Learning for LLM Reasoning from A Cross-Domain Perspective

Finding 4: RL Expanding Performance Limits is NOT a False Promise

Our findings strongly suggest that RL can teach compositional skills that are novel to the base model.

However, recent work claims that RL merely "reranks" model responses, distilling pass@k performance of the base model into pass@1.



We revisit this claim by performing analysis at varying difficulty levels.

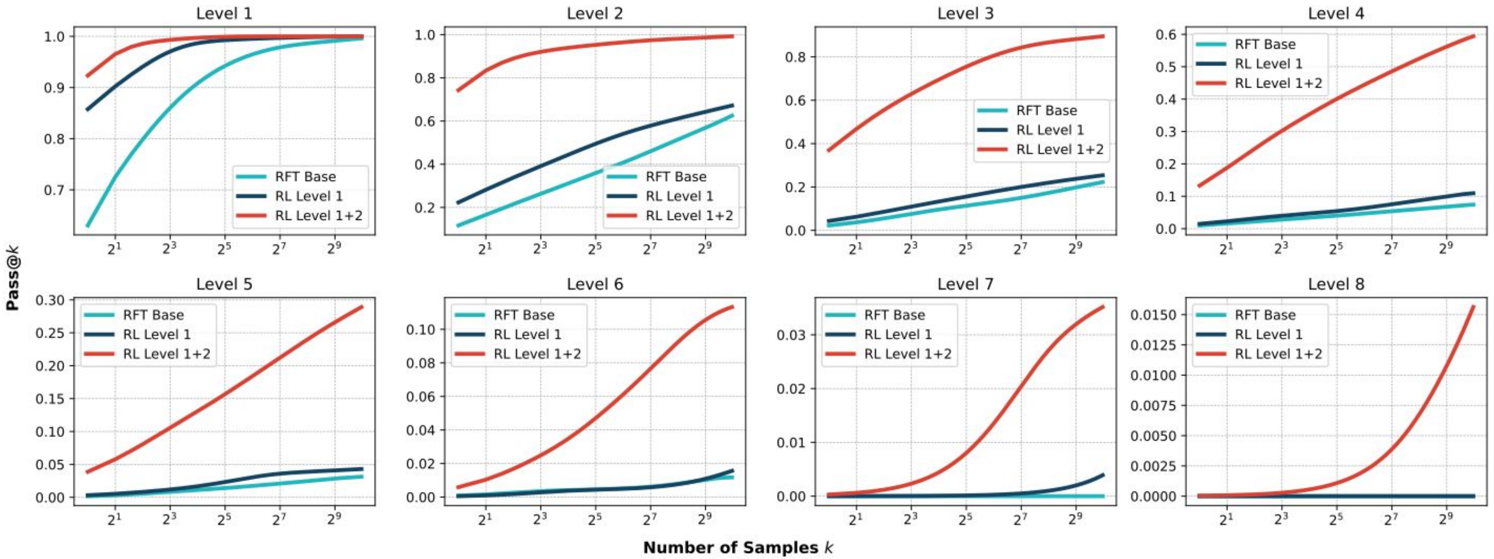
Yue et al. 2025. Does Reinforcement Learning Really Incentivize Reasoning Capacity in LLMs Beyond the Base Model?

Finding 4: RL Expanding Performance Limits is NOT a False Promise

RL Level 1 exhibits a similar trend to the RFT Base across almost all levels.

On easier problems (Levels 1 and 2) where the RFT base model already shows solving potential evidenced by high pass@k, the performance gaps between RL Level 1+2 model and the RFT model shrink as k increases, aligning with the trends observed in previous work.

However, a completely different trend is observed on more challenging compositional problems (Levels 3-6).



Finding 4: RL Expanding Performance Limits is NOT a False Promise

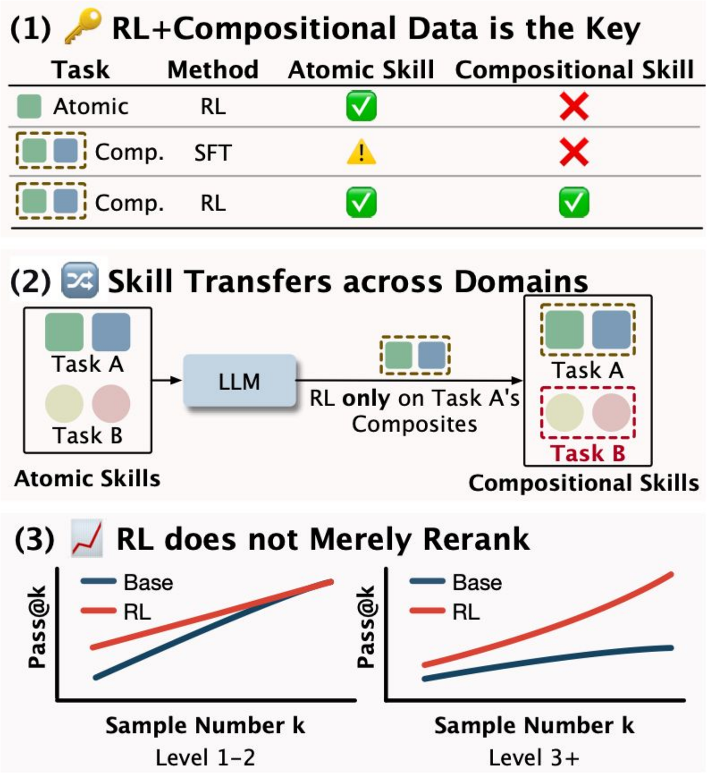
Why's the case? Two conjunctures:

- Previous work **evaluates and trains on tasks that base models already achieve high pass@k**; thus RL has little incentive to learn a new skill.

- The aggregate pass@k metrics on **mixed-difficulty benchmarks** can mask an improvement in a specific skill like composition, if other required skills remain a bottleneck.

In contrast, we provided decontaminated evaluation and fine-grained analysis on challenging problems.

Findings Review



- LLMs learn new skills by composing old ones, and both **RL** and **incentivization to composition** are required
- The learned skills **generalizes** to held-out evaluation, more difficult problems, and even a different task.
- The learning of new skills can also be reflected in pass@k when evaluation is done rigorously.

Closing Remarks

Thoughts

- Highlight the critical role of RL for its generalizability

  Is it possible to iteratively internalize the latest composition as new atomic skills, so that the model can eventually solve extremely complex problems directly?

Closing Remarks

Thoughts

- Build base models with various atomic skills

  By equipping base models with atomic skills, can we leverage the cross-task transferability of compositional RL to reduce data collection efforts on tasks where gathering RL data is costly or difficult?

Closing Remarks

Thoughts

- Closer coordination between pre-training and post-training from a skill acquisition perspective: Pre-training should optimize not only for base model performance, but for post-training potential and efficiency

Can we propose some metrics for base model to indicate post-training performance from this perspective?

Can we flexibly reallocate data collection efforts between pre-training and post-training based on where data is easier to obtain?

Thanks!

Arxiv: https://arxiv.org/abs/2509.25123

ILLINOIS