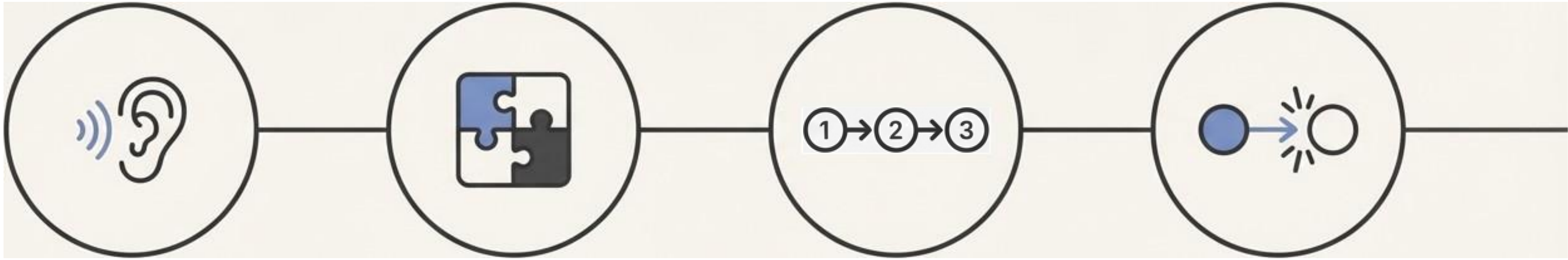
CtrlAct

Grounding LLMs to Bridge the Gap Between Instruction and Action

Qingyang Xiao, Bo Su, Ling Sun, Zhu Zhu, Thai Le
Indiana University

A Chain of Embodied Reasoning



Goal Interpretation

Translating natural language commands into grounded symbolic objectives.

Subgoal Decomposition

Inferring the necessary intermediate states and causal preconditions.

Action Sequencing

Ordering concrete, low-level operations to satisfy all preconditions.

Transition Modeling

Predicting how an action will alter the environment's state.

A failure in any single link breaks the entire chain.
We need to understand which links are weakest.

CtrlAct Framework

We evaluated three interventions to address common LLM failure modes. This allows us to disentangle errors in online reasoning from gaps in underlying knowledge.

Two open-sourced LLM models

GPT-OSS-120B (High reasoning)

Experiment setup
8 NVIDIA L40S GPUs
vLLM 0.11.0
no quantization

Qwen3-Next-80B-A3B-Thinking

Experiment setup
4 NVIDIA L40S GPUs
vLLM 0.11.0
no quantization

Intervention 1: Can Better Instructions Fix Reasoning?

An exploration of Guided Reasoning via structured prompts.

Rule-based Prompts

- Linguistic perspective
- LLM auto-generated

Two Environments

VirtualHome (Structured World): SUCCESS

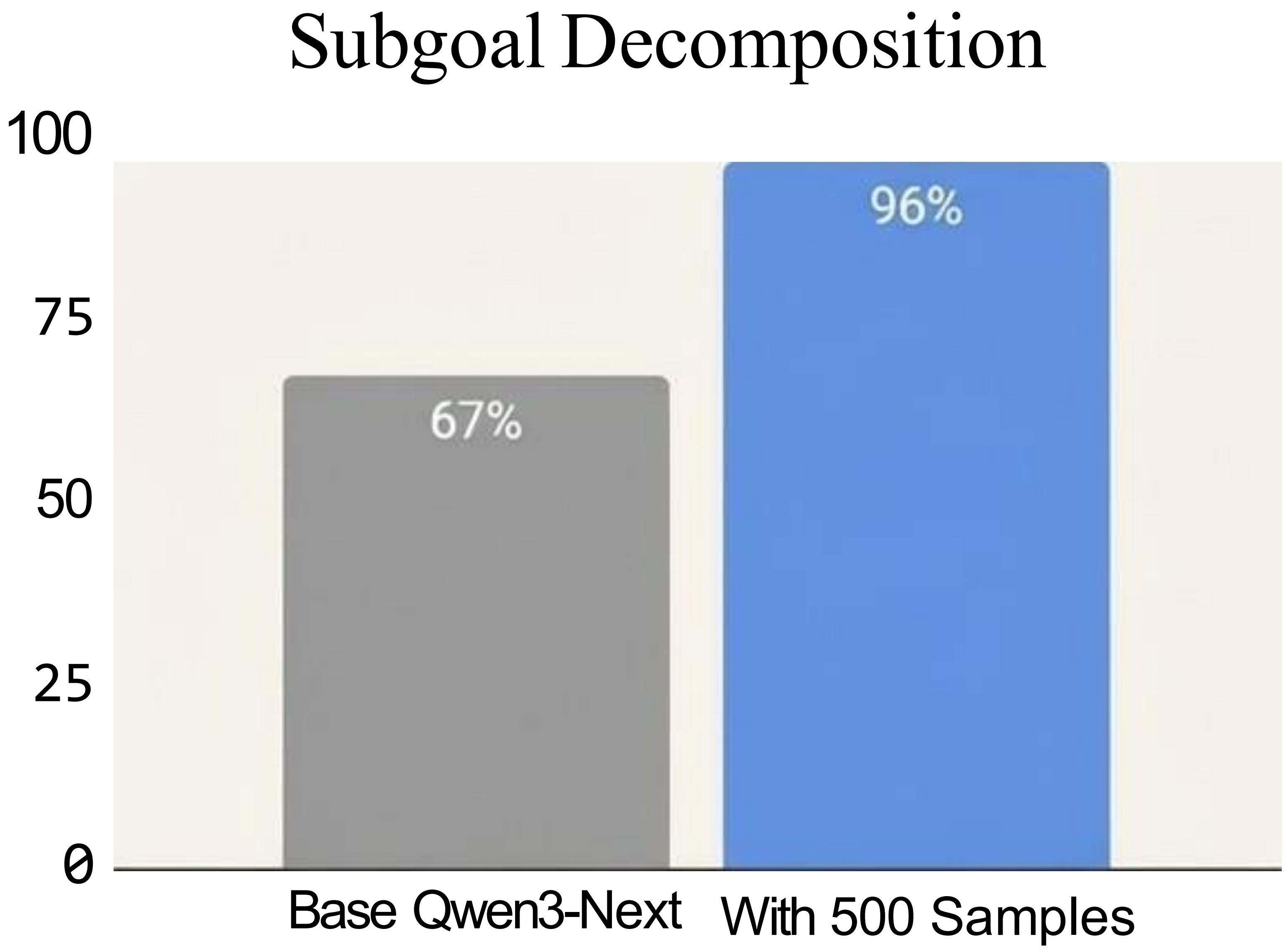
Rule-based prompt raised Goal Interpretation F1 score from 0.369 to 0.422. Structured guidance works in a regular environment.

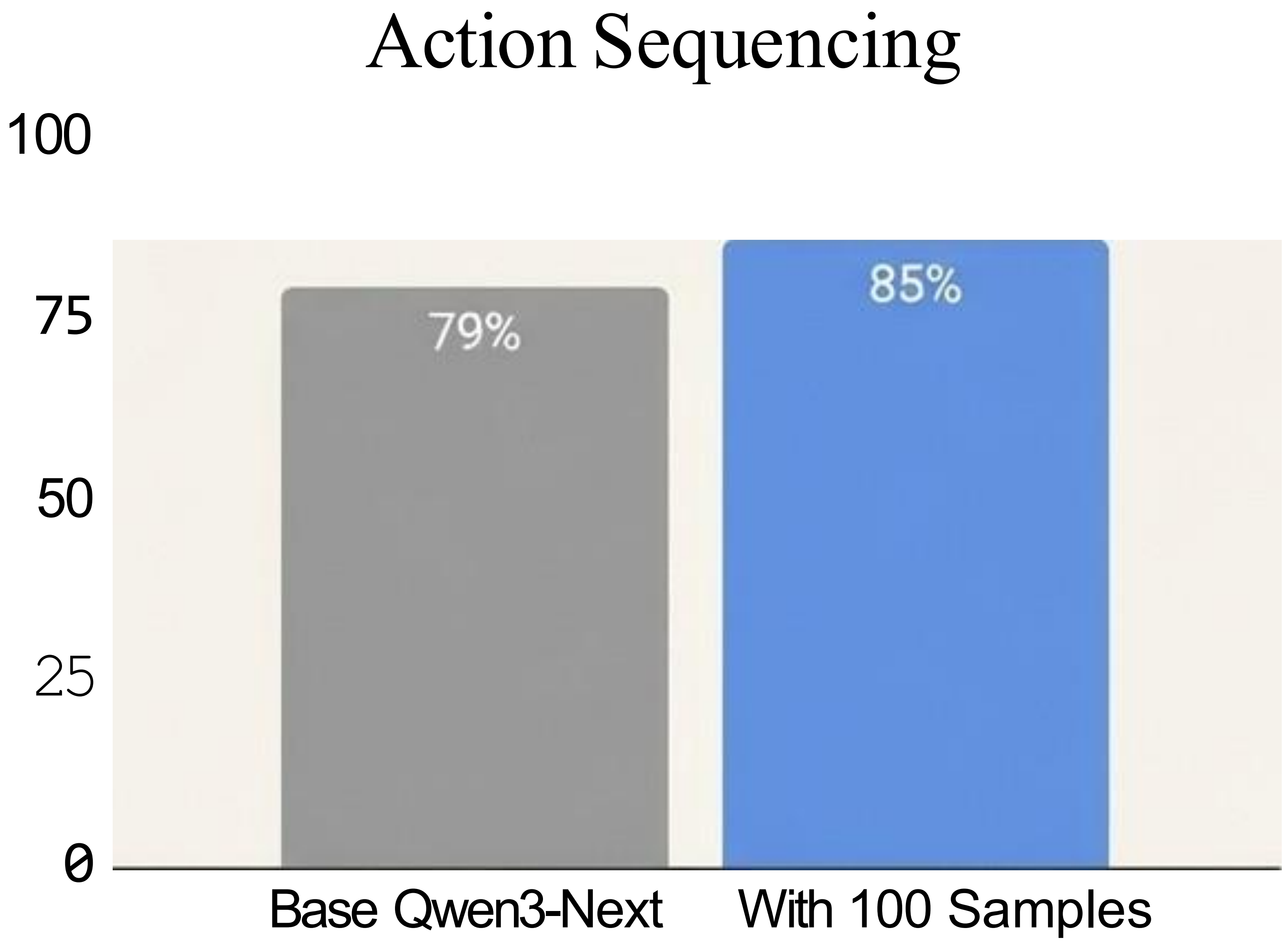
BEHAVIOR (Realistic World): FAILURE

The same methods provided no gains in planning tasks. Guidance doesn't scale to long-horizon, complex environments.

Intervention 2: Can the Model Succeed if We Let It Try More Times?

Probing model capacity with Oversampling in the BEHAVIOR environment.

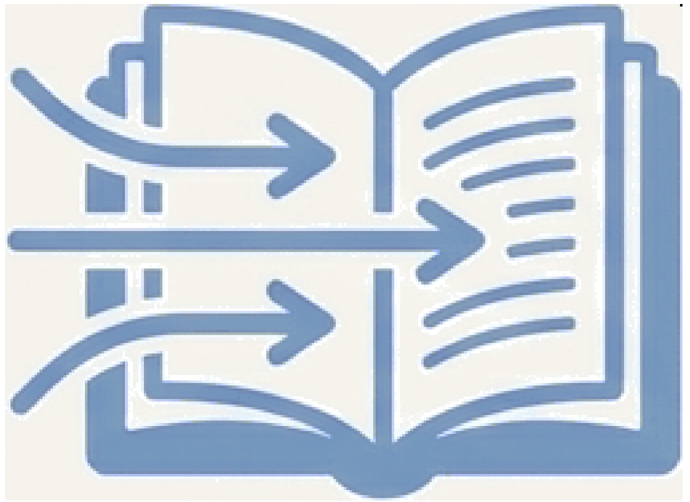




Oversampling massively helps the model figure out what steps to take, but has a much smaller effect on getting the *order* of those steps right.

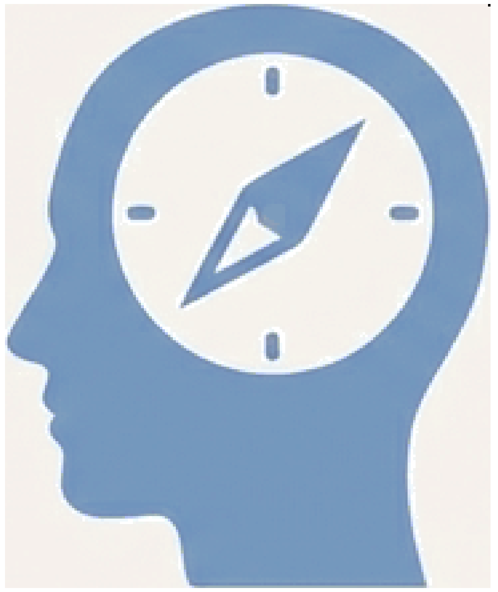
Intervention 3: Can We Reshape the Model Behavior?

A deep dive into Domain Alignment methods.



Supervised Fine-Tuning
(SFT)

Show the model perfect examples of physical cause-and-effect.



Activation Engineering
(Steering)

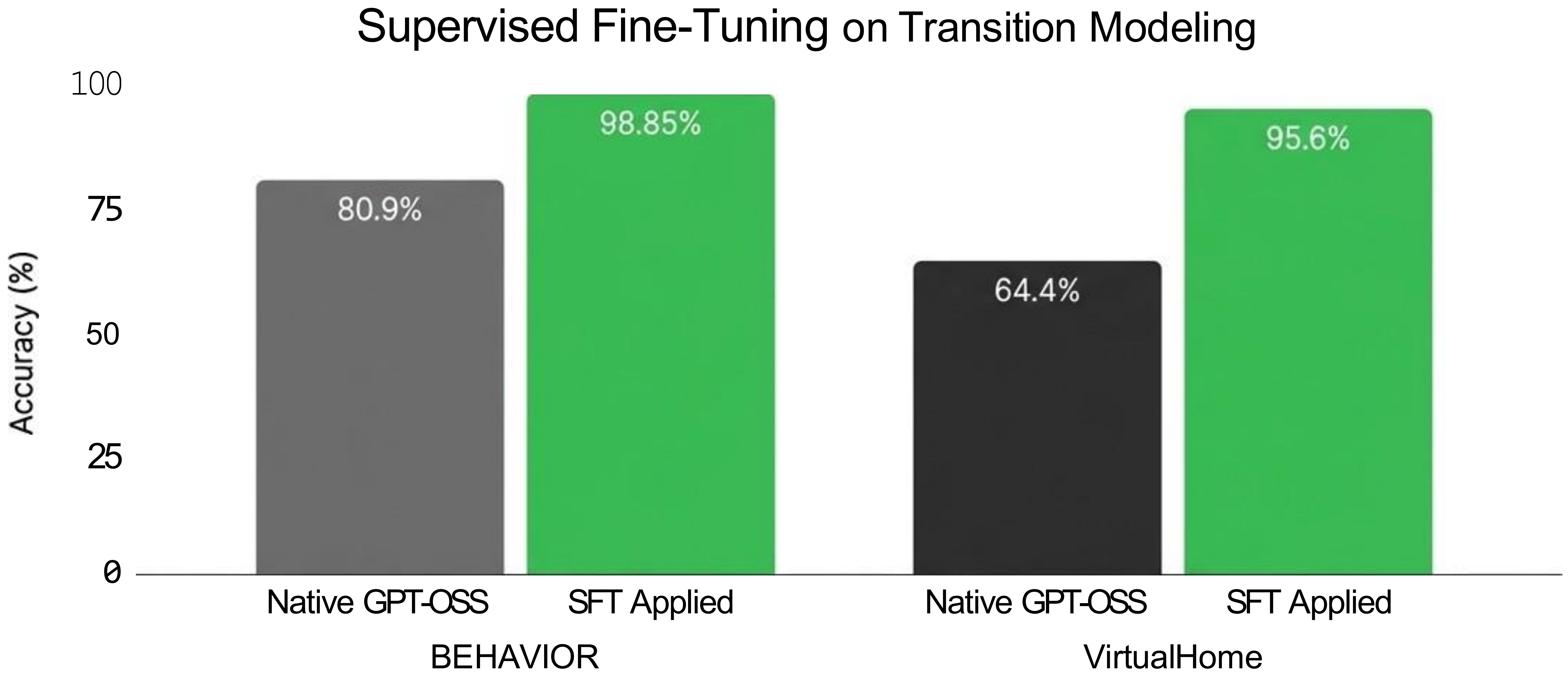
Nudge the model's internal representations toward correct reasoning paths.

Reinforcement Learning
(RL)

Let the model learn from trial-and-error with rewards for success and penalties for failure.

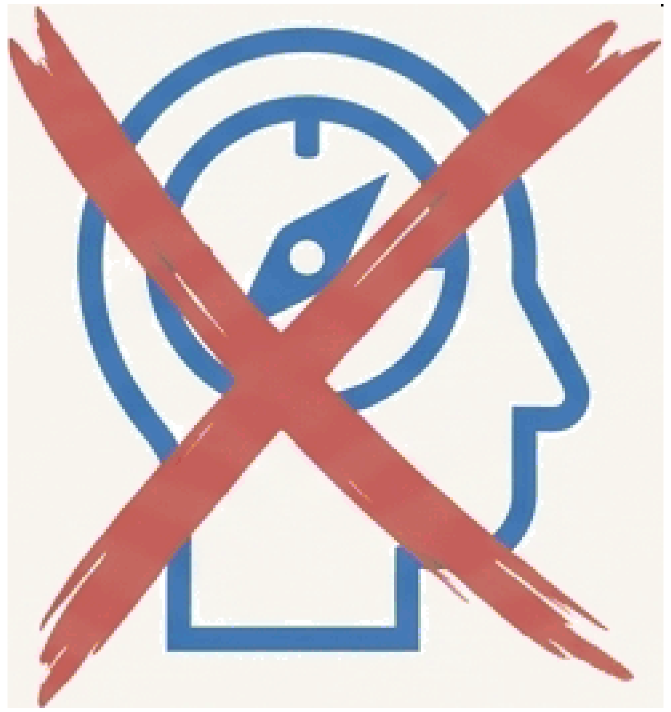
These methods aim for deeper adaptation. Which will succeed at improving complex planning?

SFT was incredibly effective. By fine-tuning on just 20% of ground-truth data, we could teach the model on fundamental rules with near-perfect accuracy.



Failed attempts

Activation Engineering (Steering)



Yielded little improvement... in some cases, performance even declined.

Destabilized ordering decisions without improving grounding.

Reinforcement Learning (RL)

Underperformed the base model on execution success, goal accuracy, and sequence validity.

Training signals were too noisy, leading the policy *away* from pretrained structure.

The methods designed to improve long-horizon planning and sequencing made things worse.

Final Results

	Behavior				VirtualHome			
	GI	**SD**	**AS**	**TM**	**GI**	**SD**	**AS**	**TM**
Native	83.2*	67†	79†	80.9*	36.9*	66.2*	71.6*	64.4*
Sys. Prompt	-	-	-	82.2*	42.2*	73.4*	72.1*	70.8*
Oversampling	-	96†	85†	-	-	-	-	-
SFT	-	-	-	98.85*	48.2*	-	-	95.6*
Final	83.2*	96†	85†	98.85*	48.2*	73.4*	72.1*	95.6*

*: GPT-OSS results; †: Qwen3-Next results. For GI and TM on BEHAVIOR, and all columns on VirtualHome, base model is GPT-OSS. For SD and AS on BEHAVIOR, base model is Qwen3-Next.

More Experiment Cases

1. Model scale

- GPT-OSS-120B vs GPT-OSS-20B
- Qwen3-Next-80B-A3B vs Qwen3-30B-A3B

2. Sampling parameters

- Q: Is Goal interpretation a translational task?
- GPT-OSS: High temperature trends to present better performance
- Qwen3: Low temperature trends to present better performance

More Experiment Cases

3. Quantization

- One example: Soap INSIDE washing machine or ONTOP washing machine
- When quantization was enabled, GPT-OSS and Qwen3-Next both failed

4. Steering and DPO

- We constructed a sample set of (positive, negative) pairs
- We observed some improvement in the training stage, but not on the evaluation stage

Takeaways

1. Native LLM behavior is competent but unreliable
2. Specific interventions are required to explore the model capacity
3. SFT excels at knowledge, not long-horizon planning
4. Guided reasoning varies by environment

Our Team

- Qingyang Xiao: PhD in Computer Science
- Bo Su: PhD in Computer Science
- Ling Sun: PhD in Linguistics & Cognitive Science
- Zhu Zhu : MS in Linguistics
- Thai Le: Assistant Professor in Computer Science

Thank you & Contact

Qingyang Xiao
CS PhD, Indiana University
Intern, Takeda

Graduate in Summer 2026

- Protein function modeling
- MS/MS spectra interpretation
- Nanobody design

Email: xiaoq@iu.edu