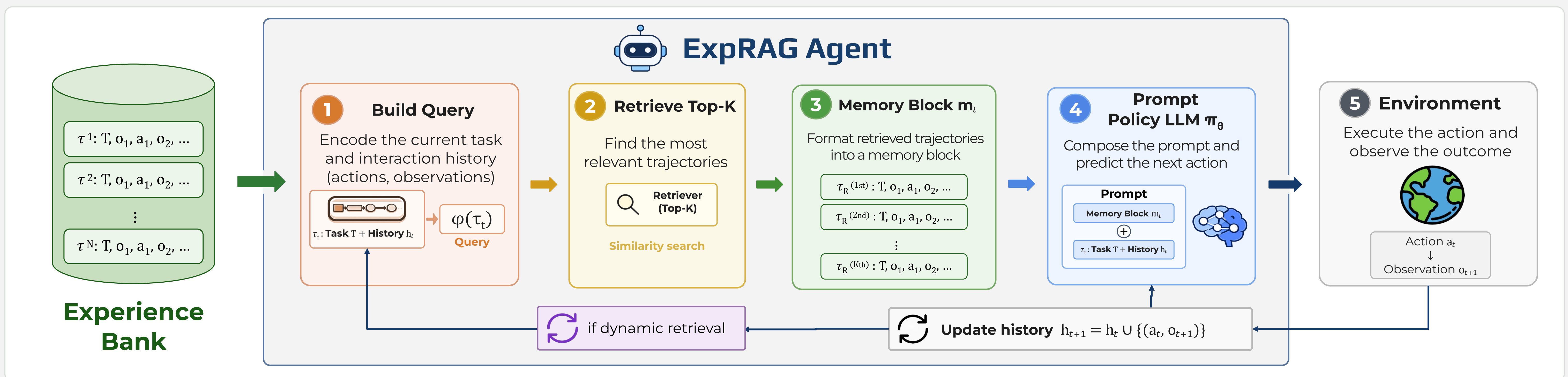
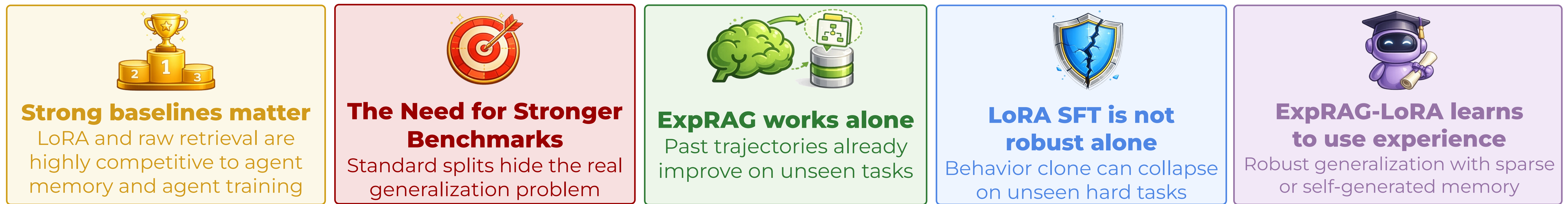


Training agents with retrieval improves OOD generalization and robustness to imperfect memory



Motivation

- LLM Agents **struggle to generalize** to unseen tasks.
- **Behavior Cloning** (SFT on Expert Trajectories) do well in in-distribution tasks but **collapse out-of-distribution**.
- **Agent memory** at inference time improve generalization but **underperform on in-domain scenarios**, as models did not learn how to use memories.

Can an LLM-based agent learn to solve new tasks by using retrieved experience in-context?

Strong baselines matter

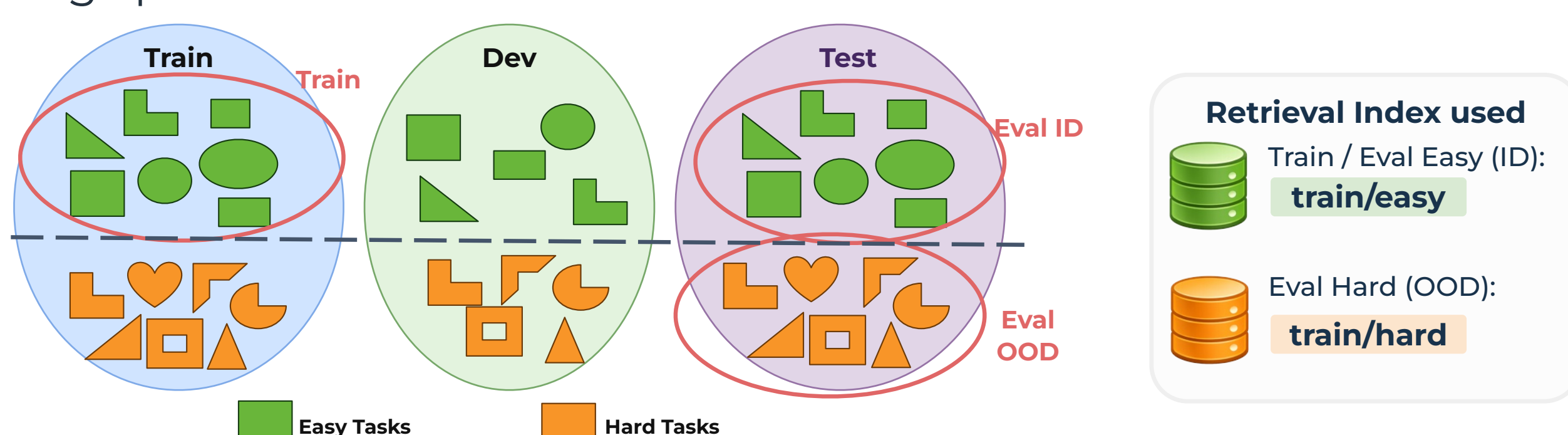
- Simple, well-tuned baselines can outperform more elaborate systems.
- **Raw retrieval trajectories beats several complex agent memories**
- Let model train for longer with LoRA SFT beats some **engineered training recipes**, almost **closing benchmark**.

Prompt-based		Training-Free Memory		Fine-tuned				
Zero shot	ReAct	A-MEM	Reflexion	Best ExpRAG (ours)	SFT w/ ReAct	SAND	Rule-based Expert	LoRA SFT (on Expert)
29.9	17.1	34.7	42.7	83.6	80.7	85.0	89.6	94.1

Cross-paper comparison on default ALFWorld valid-unseen with Qwen 2.5-7B.

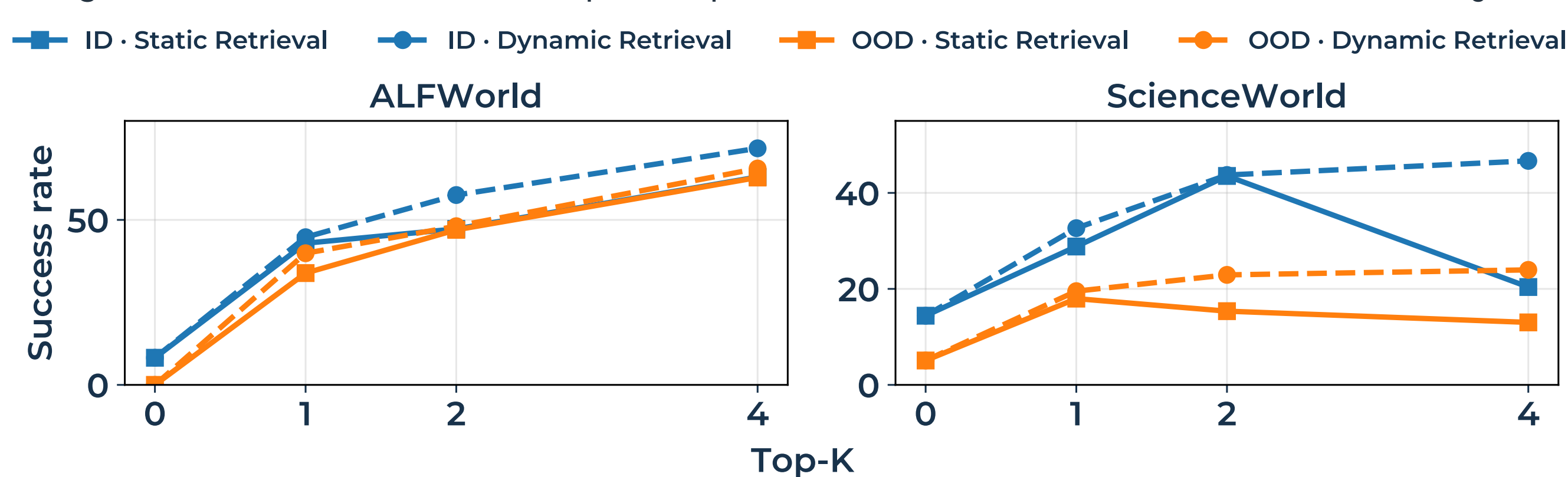
Evaluation protocol: Building OOD Benchmarks

- Benchmarks: Embodied Agents (**ALFWorld** and **ScienceWorld**).
- Train trajectories: expert success
- Index trajectories: expert (all)
- To evaluate model adaptation at test time, we partition easy/hard from existing splits:



Retrieval alone already helps

- Even without training, retrieval yields large gains over no retrieval.
- **Top-K**: more trajectories help, until long-context saturation starts to hurt.
- **Dynamic retrieval**: can improve performance, but introduces instability.



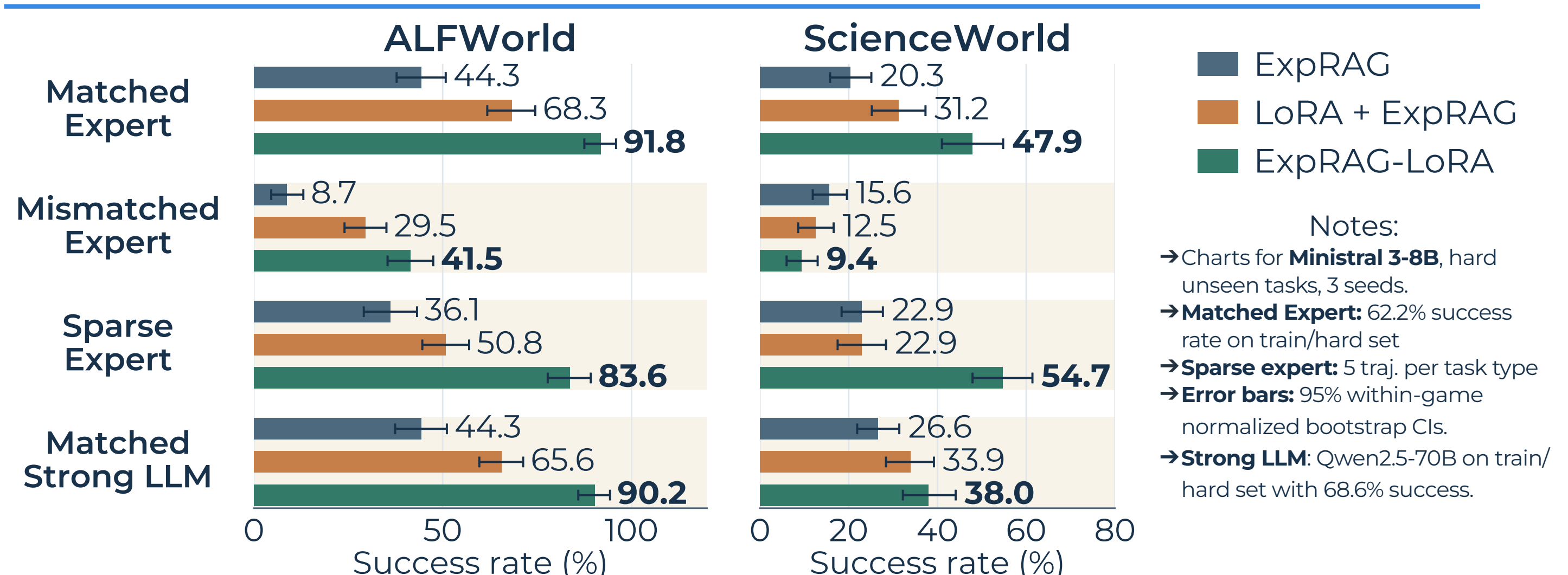
Training with retrieval gives the strongest generalization

Method	ALFWorld		ScienceWorld	
	Easy	Hard	Easy	Hard
Zero-shot	8.2	0.0	14.4	5.1
ExpRAG (inference only)	54.8	47.5	43.5	18.8
LoRA (no retrieval)	98.6	34.4	38.8	15.6
LoRA (+ inference ExpRAG)	97.3	67.2	54.1	28.1
ExpRAG-LoRA	97.3	88.5	58.8	42.2

Fine-tuning Results on Ministral 3-8B

- **Fine-tuning alone**: strong in-domain, but **weak on hard unseen tasks**.
- **Retrieval at inference helps recover OOD performance**.
- **ExpRAG-LoRA** largely delivers the **strongest OOD generalization**.

Robustness to imperfect experience



- **Learn to use, not just imitate**: **ExpRAG-LoRA** can extract useful information even from sparse or suboptimal experience.
- **Relevance >> quantity / quality**: Even limited/ imperfect experience remains useful when relevant, mismatched experience transfers less.

Turning failed attempts into useful memory



- **Learning from failures**: **ExpRAG-LoRA** turns failed rollouts into useful memory.
- **Potential for large-small LLM cascades**: The ability to reuse failures persists when combined with other memory sources, including with stronger-LLM.

Check out our paper for:

- **Code and models** released
- Full results with **four backbone models**.
- **Retrieval choices, training dynamics and efficiency analysis**
- **Practical guidance for training agents**

