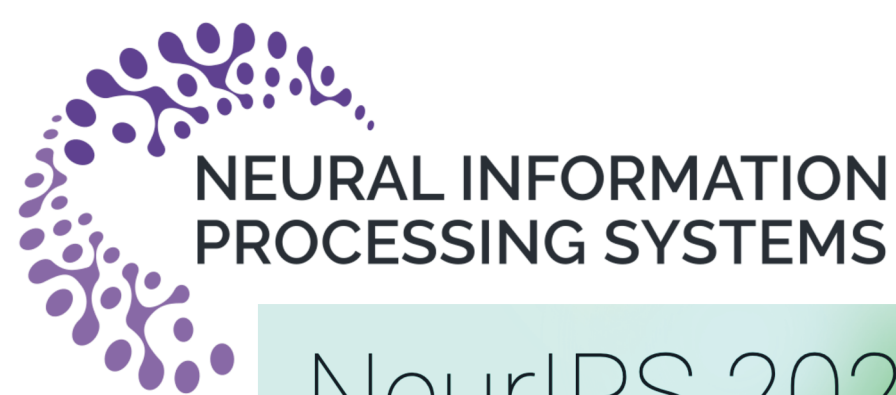


# AdaptAgent: Adapting Multimodal Web Agents with Few-Shot Learning from Human Demonstrations



NeurIPS 2024 Workshop on Adaptive Foundation Models

Gaurav Verma, Rachneet Kaur, Nishan Srishankar  
Zhen Zeng, Tucker Balch, Manuela Veloso

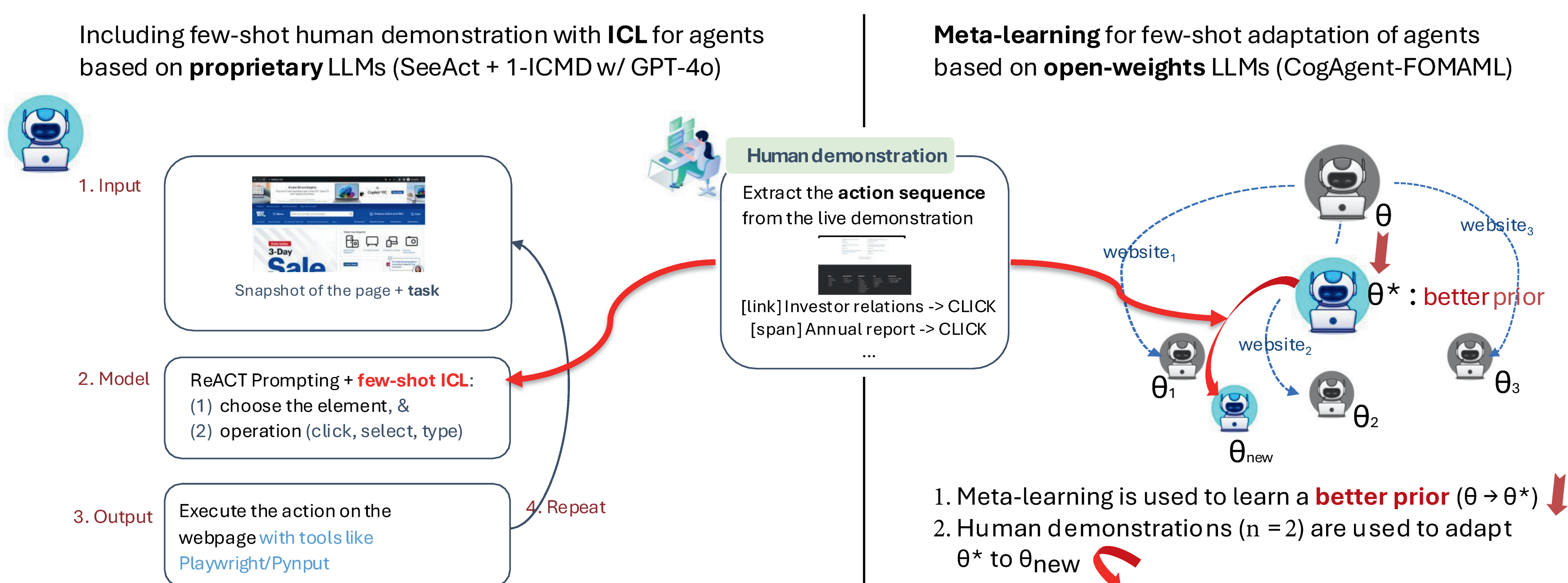


JPMORGAN CHASE & CO.  
AI Research

**How can we efficiently adapt multimodal web agents to work on unseen GUIs/webpages – including proprietary software and tasks?**

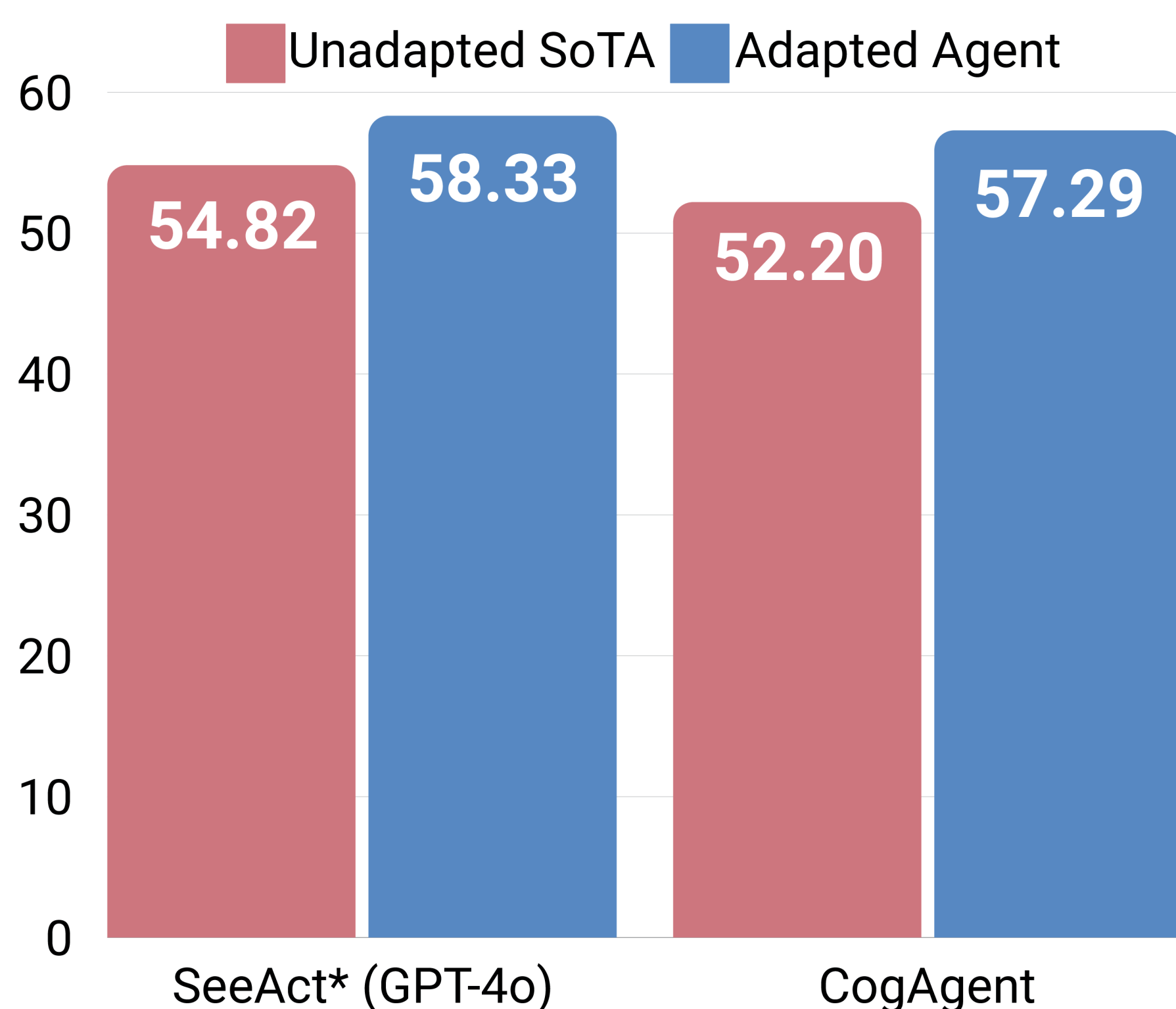
We propose the **AdaptAgent** framework!

**Upto 2 human demonstrations** to adapt MLLMs like GPT-4o (proprietary) using **in-context learning** and MLLMs like CogAgent (open-weights) using **meta-learning**.

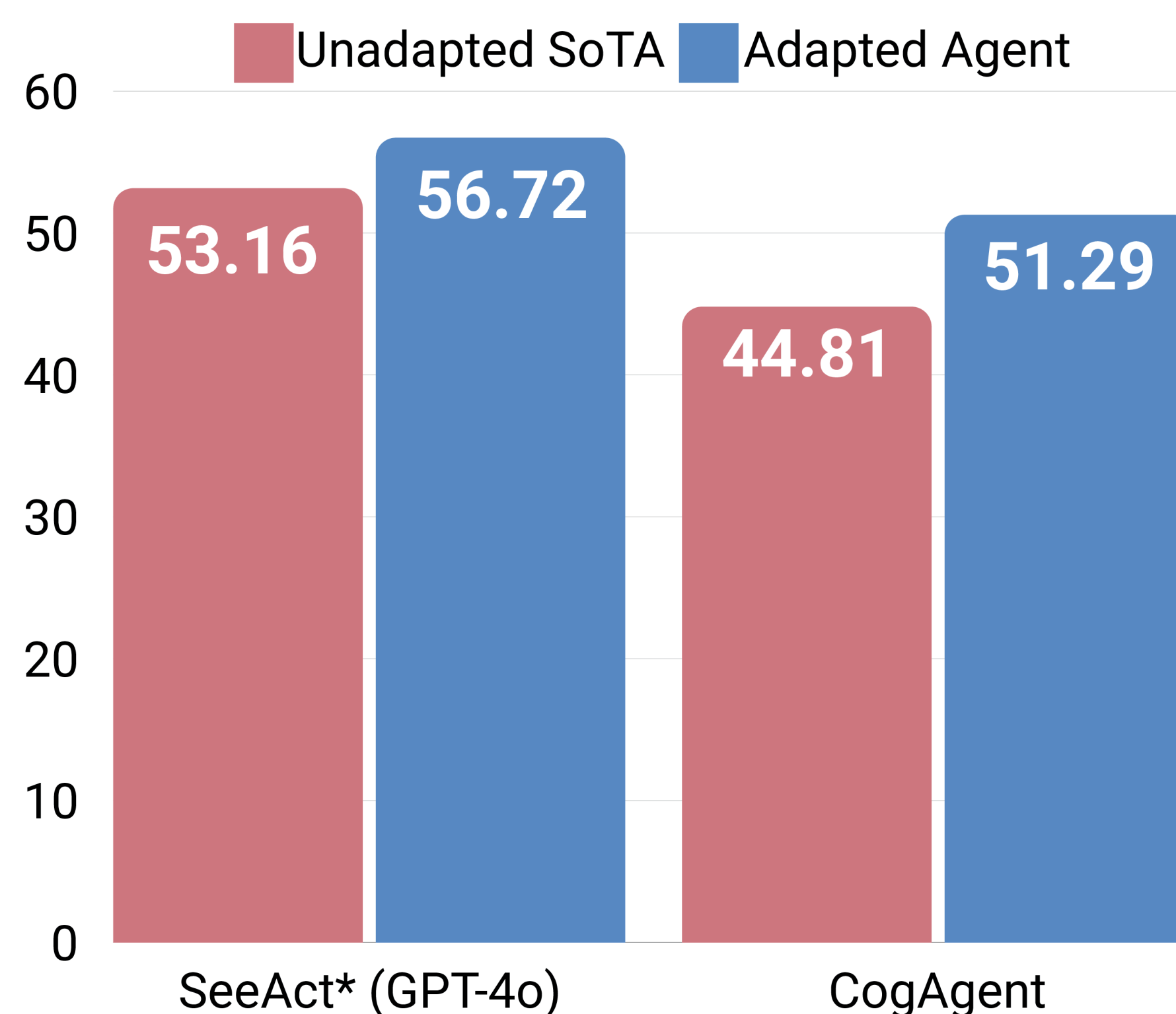


Evaluation on benchmarks like Mind2Web and VisualWebArena: our framework boosts task success rate by 3.36% to 7.21% over non-adapted state-of-the-art models, corresponding to a relative increase of 21.03% to 65.75%.

**Mind2Web Cross-Domain (Step SR)**



**VisualWebArena (Step SR)**



- Multimodal human demonstrations are more effective than text-only demonstrations
- Increasing the number of in-context learning examples helps, but with diminishing gains
- Data selection strategy during meta-learning affects the generalization of the adapted agent