ACuRL

NeurIPS 2026

Autonomous Continual Learning for Environment Adaptation of Computer-Use Agents

Tianci Xue[1], Zeyi Liao[1]*, Tianneng Shi[2]*, Zilu Wang[1], Kai Zhang[1], Dawn Song[2], Yu Su[1], Huan Sun[1]

[1] The Ohio State University [2] University of California, Berkeley

* Equal contribution







Real-world digital environments

Diversity

Millions of desktop and web applications

Dynamics

Digital environments continuously evolve over time.

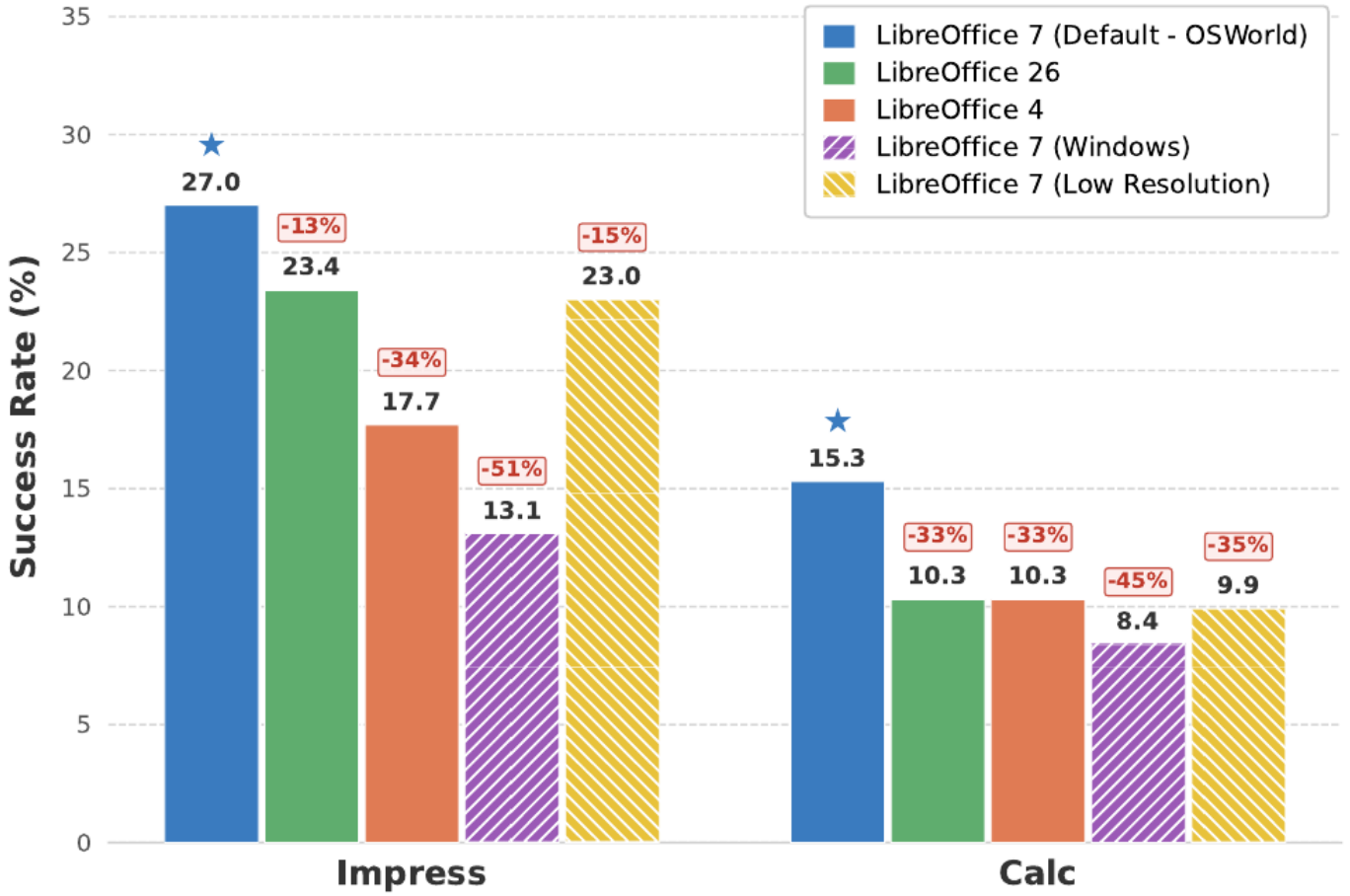


Figure 2. Impact of environmental dynamics, including software updates, platform migration, and resolution changes.

ACuRL: Autonomous Curriculum Reinforcement Learning

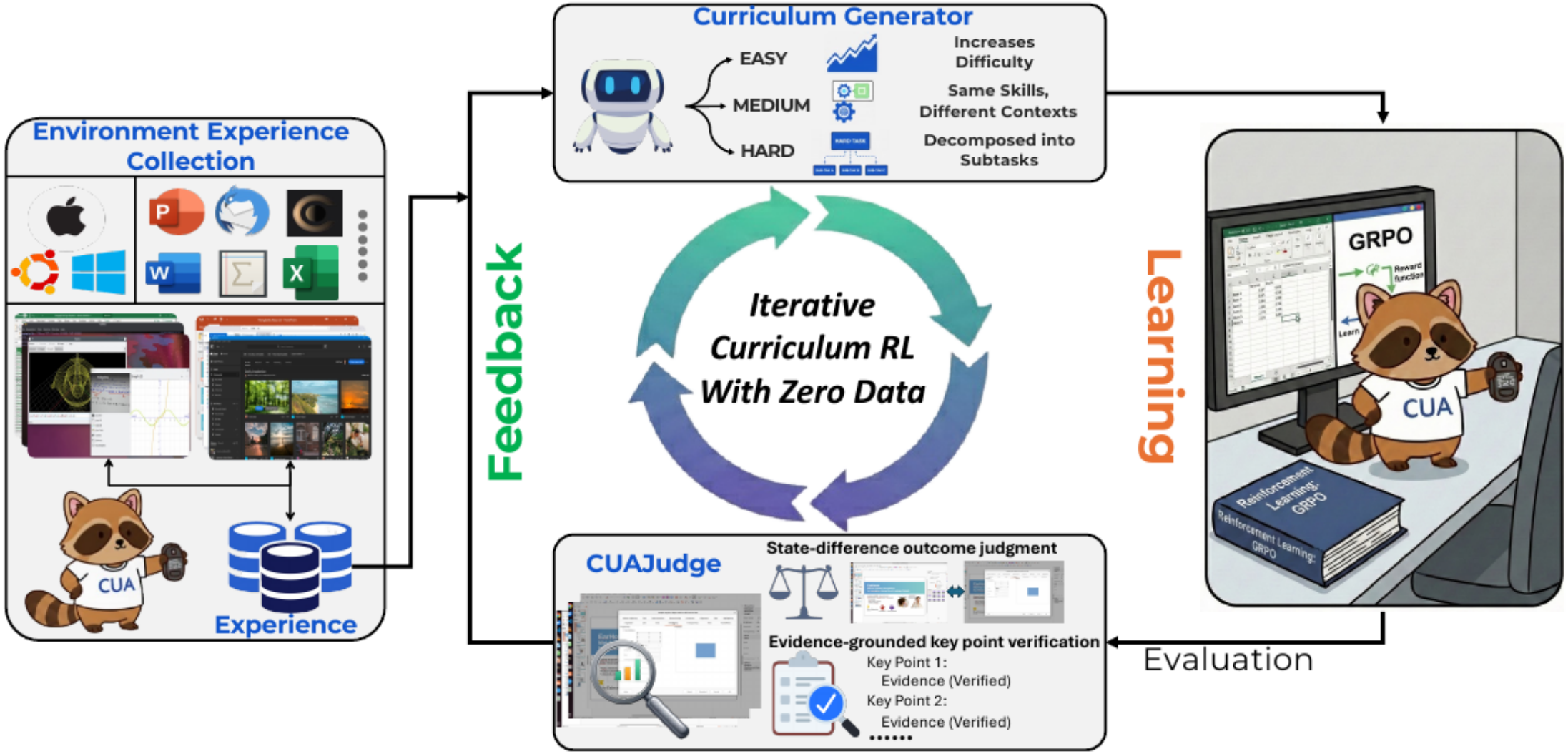


Figure 1. Overview of the ACuRL framework.

CUAJudge

Performance on several evaluator benchmarks

Evaluator	Agreement	Precision	Recall	F1
Public OSWorld Trajectories				
WebJudge	86.2	70.4	78.4	74.2
CUAJudge	**87.7**	**73.6**	**80.8**	**77.0**
AgentRewardBench				
WebJudge	81.5	73.2	48.8	58.5
CUAJudge	**83.2**	**76.6**	**53.4**	**62.9**
CUAStepBench				
WebJudge	68.0	71.9	63.5	67.4
CUAJudge	**74.8**	**76.6**	**74.5**	**75.5**

Reliability during training

Iteration	Precision	Recall	Agreement
Iter 1	98.7	97.4	97.8
Iter 2	93.2	88.5	89.9
Iter 3	94.2	92.9	93.5
Overall	**95.4**	**92.9**	**93.7**

CUAJudge achieves high agreement with human judgments across diverse scenarios.

Intra-environment continual learning

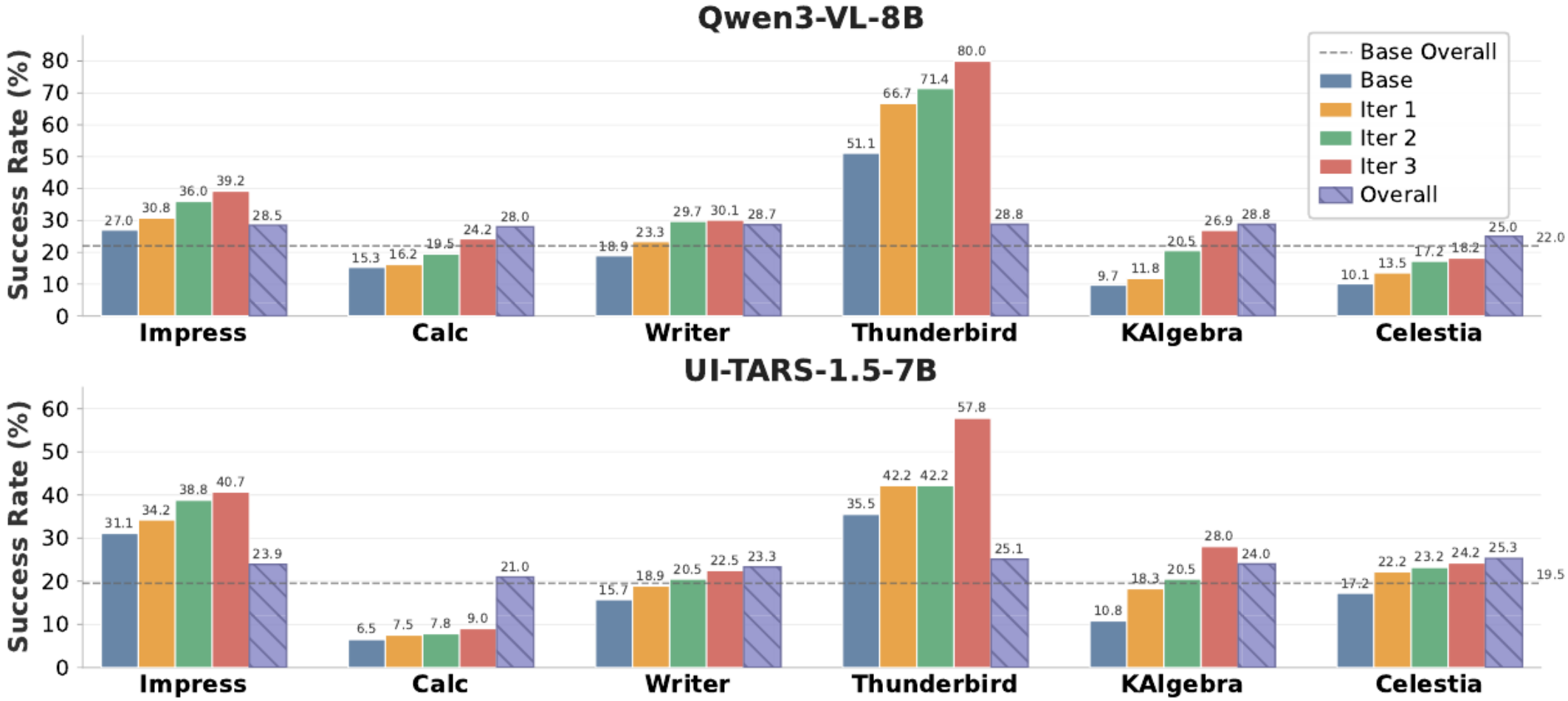
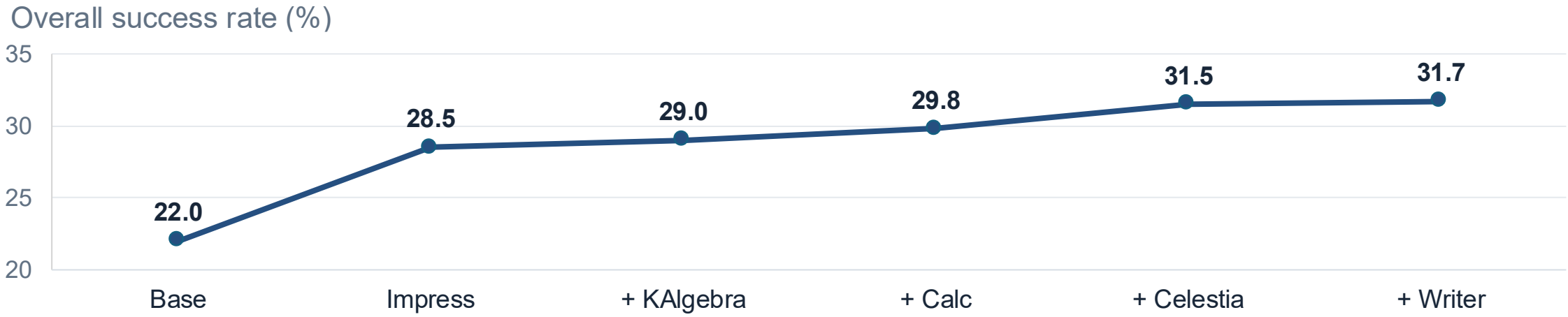


Figure 3. Average success rates of base agents and agents after intra-environment continual learning with ACuRL across six environments.

Performance improves progressively across training iterations

Cross-environment CL without significant forgetting

Impress → KAlgebra → Calc → Celestia → Writer



Backward transfer ↑

$$\mathrm{BWT} = \frac{1}{N-1}\sum_{j=1}^{N-1}\left(R_{N,j} - R_{j,j}\right),$$

+0.58%

Forward transfer ↑

$$\mathrm{FWT} = \frac{1}{N-1}\sum_{j=2}^{N}\left(R_{j-1,j} - b_j\right),$$

+3.95%

Forgetting ↓

$$\mathrm{Forgetting} = \frac{1}{N-1}\sum_{j=1}^{N-1}\left(\max_{i\in\{j,\ldots,N\}} R_{i,j} - R_{N,j}\right).$$

1.00%

$R_{i,j}$: performance on environment j after learning through environment i. b_j: base-agent performance on environment j.

Adaptation to environment changes

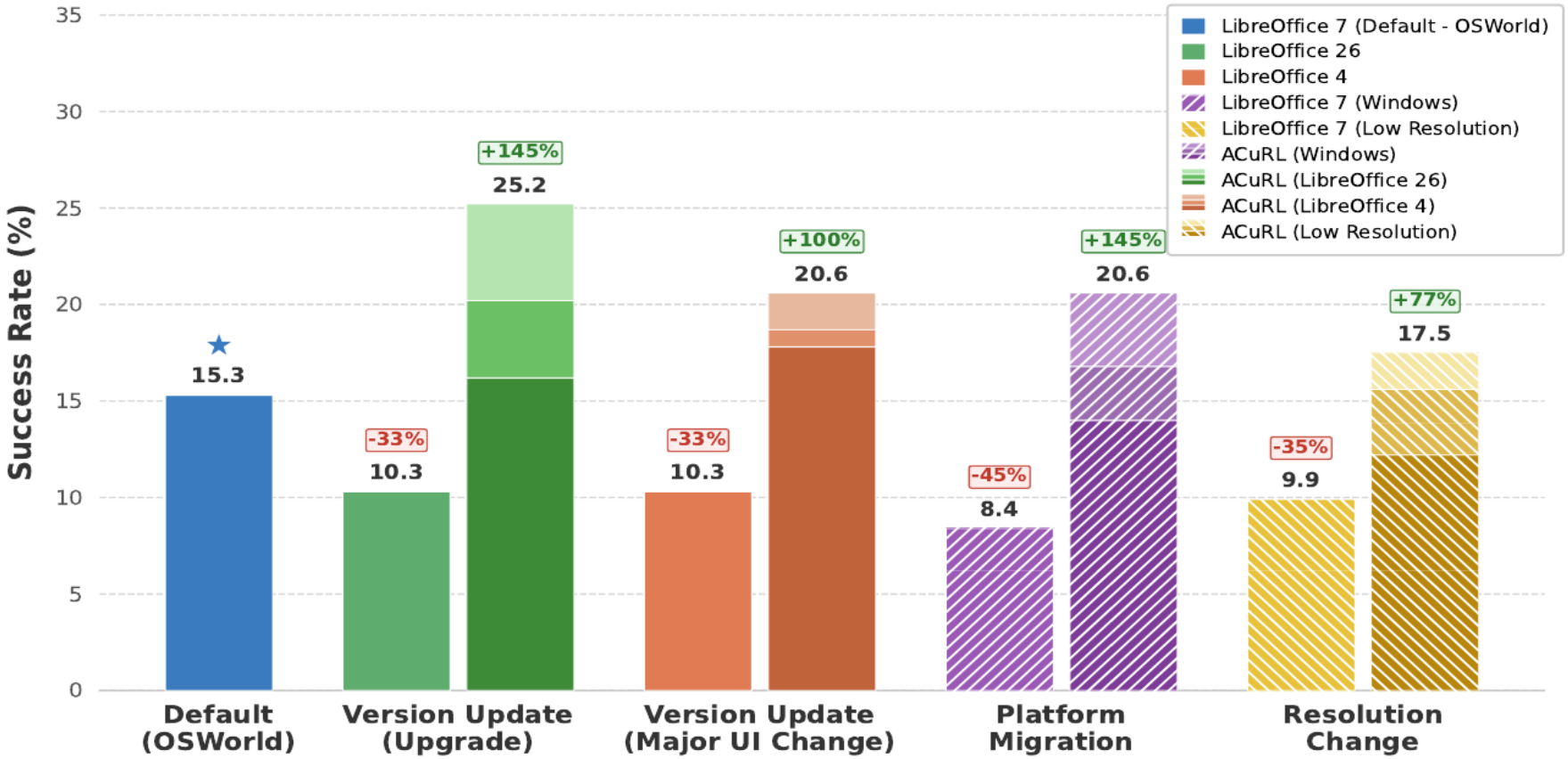
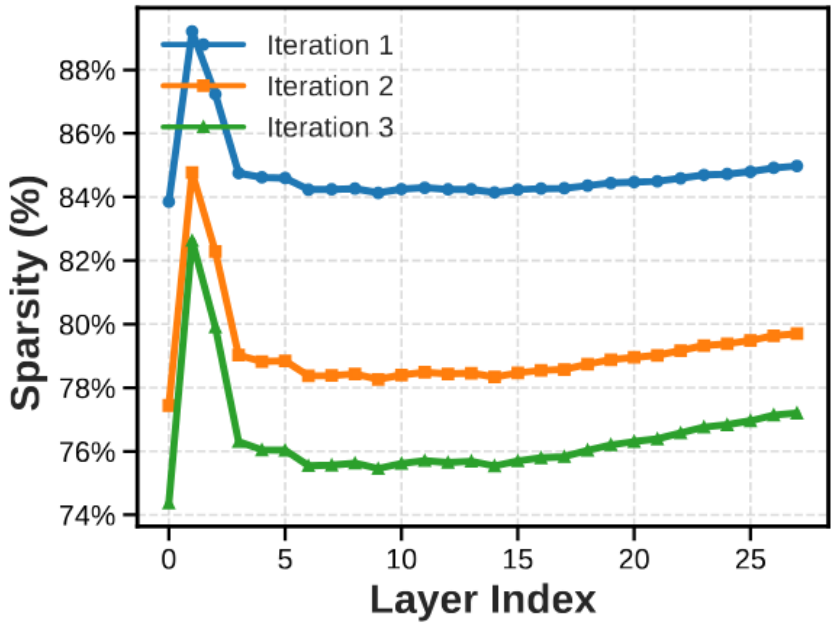


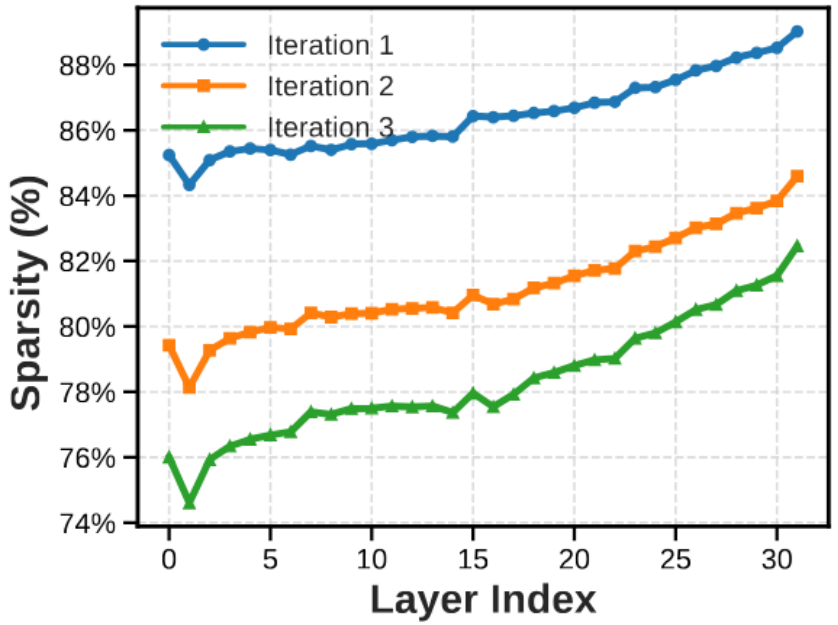
Figure 4. Performance of ACuRL under environment changes, including software updates, platform migration, and resolution changes. Results are reported on LibreOffice Calc.

ACuRL effectively mitigates performance degradation under environment changes, achieving up to 145% relative improvement.

Update sparsity during continual learning



(a) LLM backbone



(b) Vision encoder

$$\text{UpdateSparsity}_{\varepsilon}(\theta_0, \theta_1) = 1 - \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(\left|\theta_1^i - \theta_0^i\right| > \varepsilon\right), \quad \varepsilon = 10^{-5}$$

Figure 5. Sparsity of parameter updates across RL iterations, averaged over six environments.

Approximately 80% of parameters remain effectively unchanged