Salesforce AI Research

NeurIPS 2026

CLIFT

Conformal Self-Verification for Web Agent Training and Test-Time Scaling

Yifan Zhang, Yutong Dai, Viraj Prabhu,

Zhiyuan Hu, Ran Xu, Zeyuan Chen

Salesforce AI Research



Verification limits web-agent learning

Binary task success

- Cheap and deployable
- Sparse credit assignment
- Identical outcomes give zero group advantage

Frontier LLM judge

- Rich feedback
- Expensive during training
- Cannot assume judge access at deployment

KEY QUESTION

Can the agent reuse training-judge feedback at deployment?

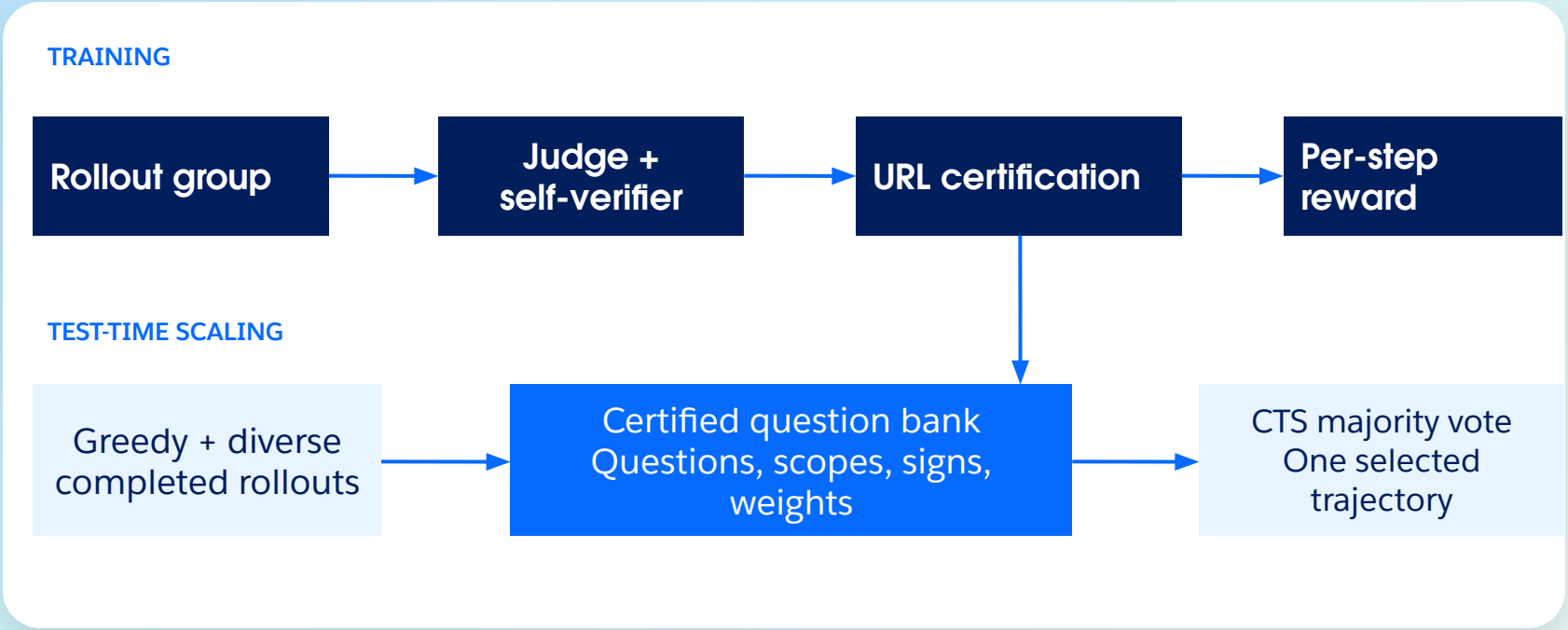
CLIFT turns judge feedback into a reusable verification bank

URL-local evidence for browser actions

- Vanilla GRPO broadcasts one trajectory reward across actions.
- CLIFT asks YES/NO questions at each visited URL.
- Actions inherit evidence from questions certified for that URL.



One Certified Bank, Two Uses



Training
Certified per-step reward

Shared artifact
Updated in training, frozen in CTS

Deployment
Selection without an external judge

What the verification bank contains

1. **Relevant content w = +0.54**
   Did the agent reach a page containing task-relevant information? (WAI)

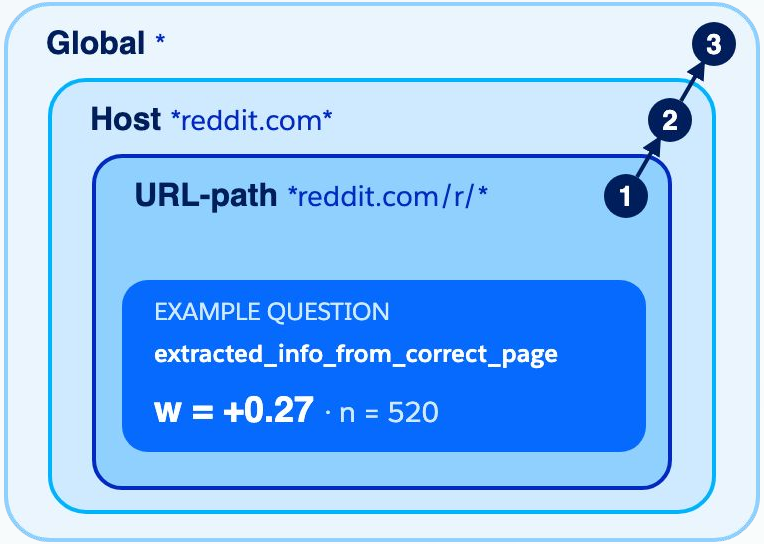
2. **Grounded final answer w = +0.27**
   Did the final answer come from the page containing the requested information? (WAI)

3. **Search followed by action w = +0.20**
   Did the agent submit a search and then act on its results? (VWA)

4. **A reusable evidence schema**
   Every question stores its URL scope, polarity and certified trust weight.

Positive weights mark progress. Negative weights mark failure evidence.

Compositional Conformal Certifier



Query the most-specific certified tier

URL-path → host → global

1 Question

YES/NO check, URL scope and polarity
Base model answers with task adapter detached

2 Certify

Track polarity-aware agreement with the judge separately at global, host and URL-path tiers

Trust depends on measured separation and reliability

3 Weight

$$w = \sigma \cdot |\text{lift}| \cdot \rho \in [-1, +1]$$

ρ discounts weak or unstable evidence

CEGAR adds 3–5 questions every 60 steps. New questions start at zero trust.
ACI tracks miscoverage. It is a diagnostic outside the trust-weight equation.

Per-step reward and URL-local credit

Certified verifier evidence can only increase the judge reward

1

Judge

One judge call scores G = 8 rollouts
Trajectory and per-step feedback

2

Self-verify

Base model answers questions per rollout and URL state
$R^{VQ}(u) \in [-1, +1]$

3

Blend

$\tilde{r} = r + \beta \max(0, v)$
$\beta = 0.15$

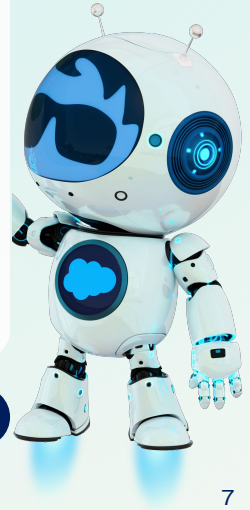
r: judge reward
v: verifier score

Reward ≥ judge

4

URL-stratified advantage

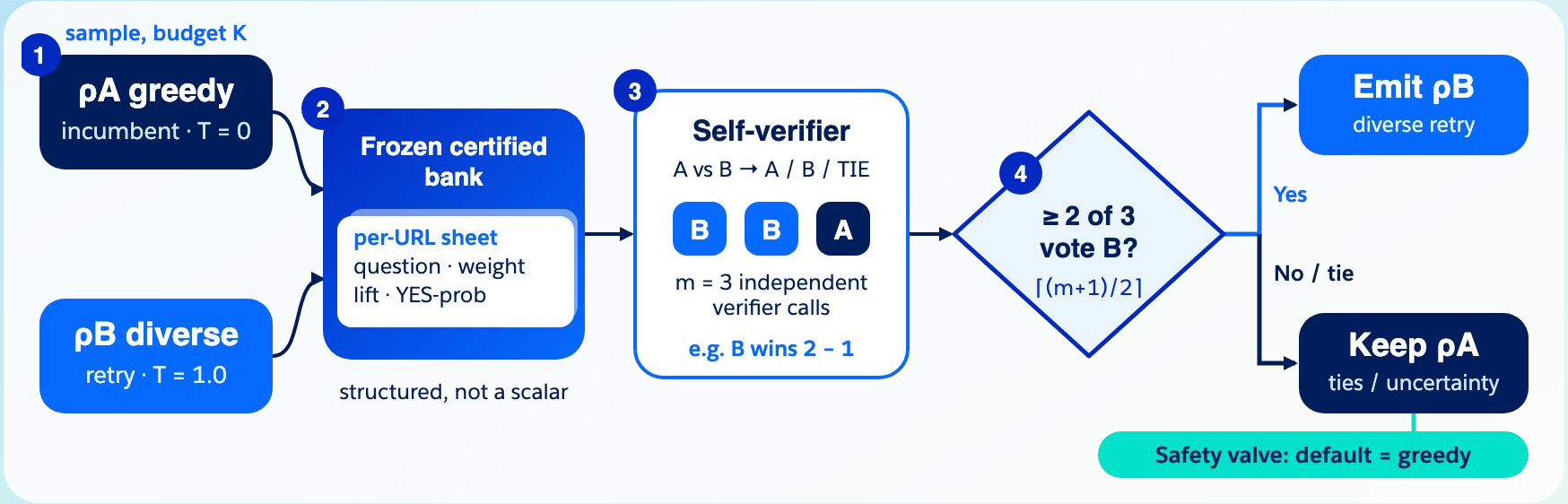
Discount returns ($\gamma = 0.95$)
Normalize by group, step and URL
Apply to tool-call tokens



Judge (Claude Opus 4.6) is training-only; never called by CTS

Conformal Trajectory Selection (CTS)

Compare completed rollouts using a frozen bank and the base self-verifier



Conservative default
At least 2 of 3 votes to swap
Otherwise keep incumbent

K > 2
Tournament over rollouts
One selected trajectory

Judge-free CTS
K counts rollouts
m = 3 verifier votes

Three complementary evaluation settings

Transfer

VWA

VisualWebArena

Bank learned while training Gemma-4

EVAL

Transfer bank to frozen GPT-5.5 (Classifieds, Shopping, Reddit)

Train

WAI

WebArena Infinity

Train Gemma-4 31B + LoRA (rank 8); programmatic per-app verifiers

EVAL

In-domain + public 9-app board

Live web

OM2W

Online Mind2Web

No policy training
120 labeled rollouts recalibrate the bank

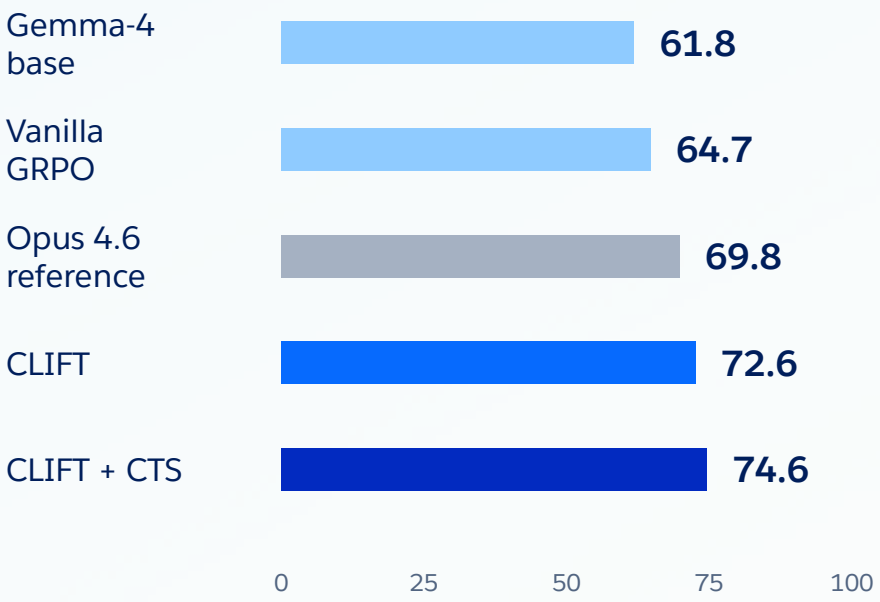
EVAL

300 live-web tasks
180 outside calibration

1,886 WAI + 1,272 VWA synthesized training tasks 16 LoRA shards

WAI: 74.6% success across nine apps

WAI success rate (%)



74.6%

Gemma-4 31B + CLIFT + CTS

9-app success · K=3
+12.8 pp vs base
7/9 apps won

Gemini 3 Flash + BU: 70.1%
Highest among the systems compared

Where the gain comes from

+10.8 pp
Training over base
61.8% to 72.6%

+2.0 pp
CTS over CLIFT
72.6% to 74.6%

WAI gains primarily come from training

Same browser-use backbone and training judge. Solve steps average successful tasks.

Variant	Change	SR	Δ vs base	Avg solve steps
Gemma-4 base	no RL, greedy	61.8	–	15.4
Vanilla GRPO	scalar judge reward	64.7	+2.9	18.5
CLIFT	certified step reward + URL advantages	**72.6**	+10.8	16.0
CLIFT + CTS	conservative selection	**74.6**	+12.8	15.5
Opus 4.6	closed reference	69.8	+8.0	–

1

CLIFT gains 7.9 pp over GRPO
Average solve steps: 16.0 vs 18.5

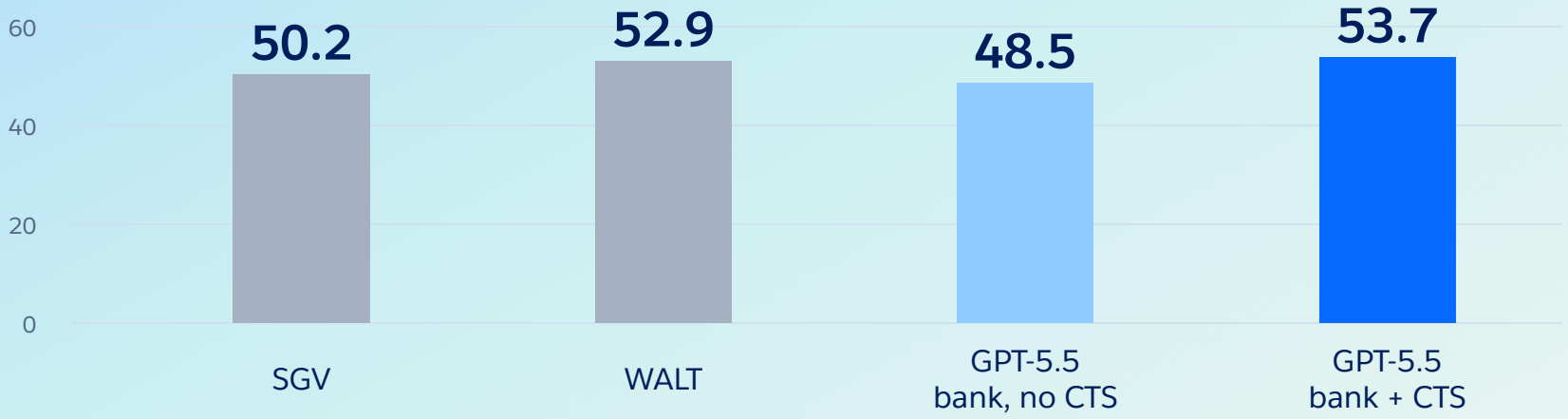
2

CTS adds another 2.0 pp
5 apps improve, 4 tie, 0 regress

17.8% swap rate

VWA: 53.7% with a frozen GPT-5.5 executor

Weighted success rate (%) under the canonical VisualWebArena harness



1 Closed executor stays frozen
48.5% to 53.7% with CTS
Rollout budget K = 4

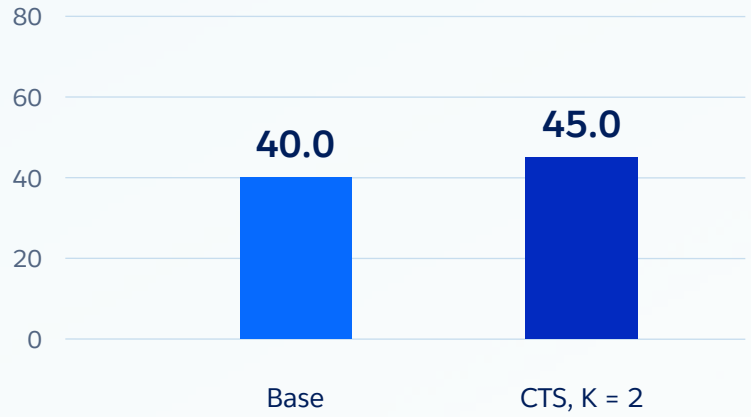
2 Open-backbone training
Gemma-4: 30.9% base, 34.4% CLIFT
35.6% with CTS at K = 2

Camera-ready Table 4

OM2W: bank transfer without policy training

Gemma-4 on live-web tasks

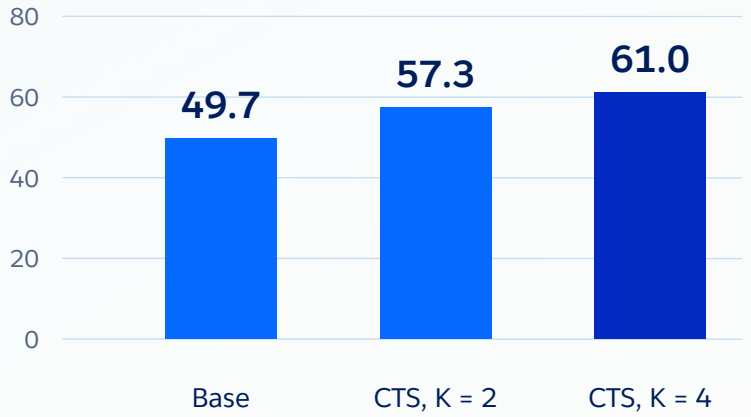
Upstream weights, no OM2W LoRA



+5.0 percentage points
Canonical judge on 300 live-web tasks

GPT-5.5 high effort

Bank translated, expanded and recalibrated



+11.3 percentage points at K = 4
Held-out 180 tasks: 45.6% to 60.0%

120 labeled rollouts calibrate the bank. No reported policy trains on OM2W.

VWA: selection gains on identical candidates

Every selector receives the same four completed GPT-5.5 rollouts per task. Pooled n = 806.

Selector	SR (%)	n	Δ vs anchor	Interpretation
Greedy anchor	45.4	806	0.0	First greedy
Random selection	45.2	806	−0.2	Uniform pick
CTS	47.6	806	+2.2	One output
Oracle union	56.1	806	+10.7	Upper bound

1

CTS vs anchor: +2.2 pp
26 gains and 8 losses
Paired p = 0.003

2

Bank-free GPT-5.5 vote
On 444 tasks, +1.1 pp for CTS
is not detectable (p = 0.46).

Camera-ready Table 10

Case study: CTS recovers a Gmail task

1

Task: profile and contact labels

Update the profile name, create the Priority label, and apply it to Family contacts.

2

Greedy incumbent: 87 steps, failed

The greedy trajectory fails to complete the requested workflow.

3

Diverse retry: 25 steps, passed

The retry completes the required profile update and contact-label changes.

4

CTS selects the successful retry

The frozen bank and comparative self-verifier choose the completed retry.

Qualitative example from a resettable WAI Gmail sandbox. Camera-ready Appendix A.9.

Conclusions

1

Certified feedback improves policy training

CLIFT raises WAI success from 61.8% to 72.6% before test-time selection.

2

The bank enables judge-free test-time selection

CTS adds 2.0 percentage points on WAI and emits one selected trajectory per task.

3

Verification knowledge transfers

VWA transfer reaches 53.7%. Recalibrated OM2W transfer reaches 61.0%.

CLIFT turns costly judge feedback into reusable verification knowledge