40th Conference on Neural Information Processing Systems

Sydney (Main)
Dec 6 – Dec 12, 2026

NeurIPS 2026 Poster

HyperSkill: Training-Free Omnimodal GRPO via Hypergraph-Indexed Skill-Library Evolution

Haoran Luo[1†], **Shangkai Lin**[1], **Jinyang Wu**[2], **Xinliang Zhou**[3*], **Luu Anh Tuan**[1,4*]

[1] Nanyang Technological University [2] Tsinghua University
[3] Stanford University [4] VinUniversity
† *Project Lead* * *Corresponding Authors*

haoran.luo@ntu.edu.sg, xinliang.zhou@stanford.edu, anhtuan.luu@ntu.edu.sg

CONTENTS

HyperSkill: Training-Free Omnimodal GRPO via Hypergraph-Indexed Skill-Library Evolution



NEURAL INFORMATION PROCESSING SYSTEMS

1 Background of Study

2 Contributions

3 Methodology

4 Experimental Results

1

PART

Background of Study

NEURAL INFORMATION PROCESSING SYSTEMS

HyperSkill: Freeze the Backbone, Evolve the Skills

GRPO-style reinforcement learning is now the main way to improve LLM reasoning, but improving an agent on omnimodal tasks (web search, long-document reasoning, audio-video QA) carries a per-domain parameter-update cost.

Parameter updates are **expensive, forgetful, and opaque**: they risk catastrophic forgetting and bury improvements in weight deltas.

Training-free skill libraries keep the backbone frozen and grow a textual experience store — but still compare whole trajectories, admit skills by co-occurrence, and store them flat.

HyperSkill keeps the backbone frozen and replaces gradient updates with auditable edits to a Hypergraph-Indexed Skill-Library.

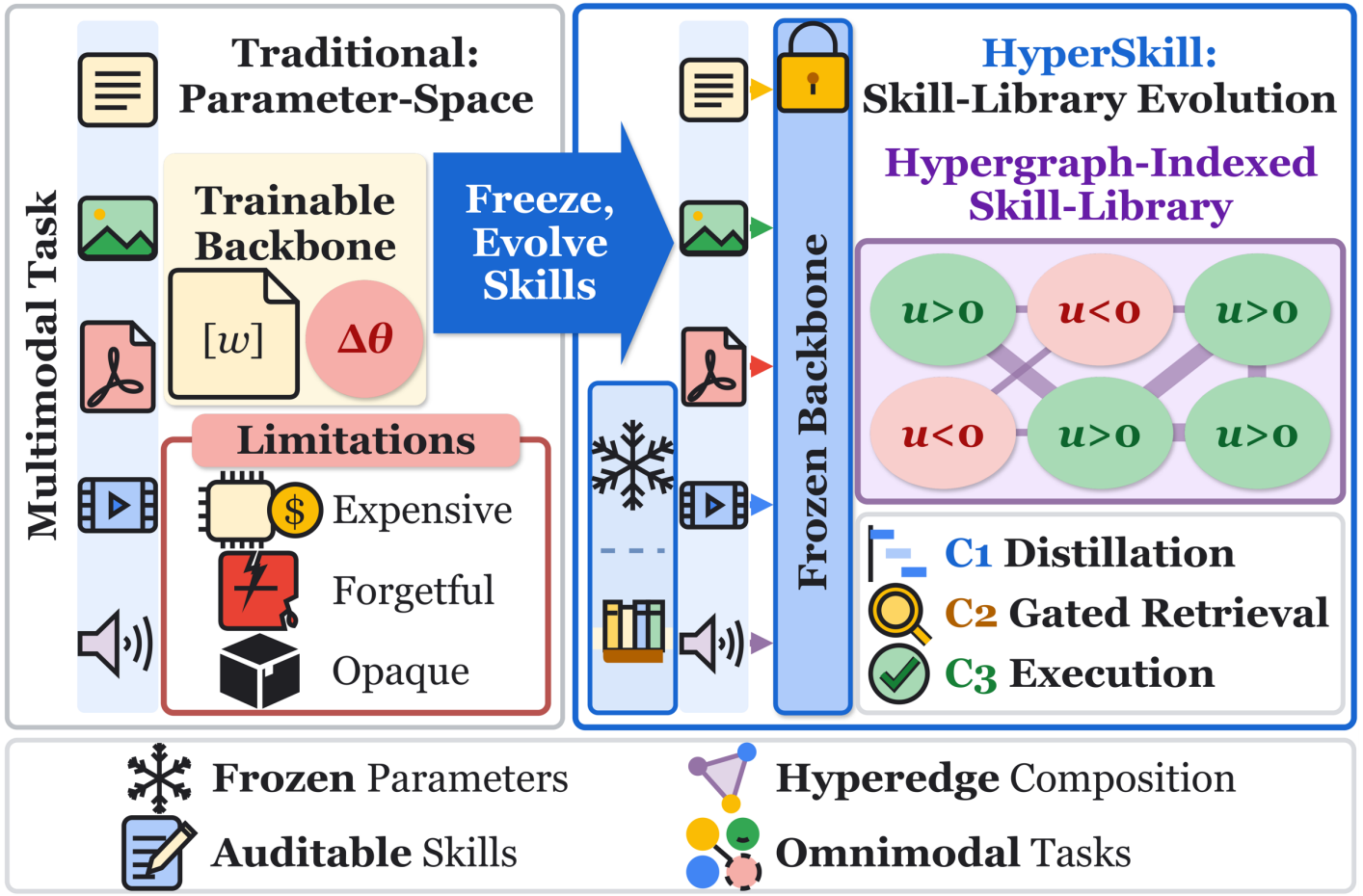


Figure 1: Parameter-space GRPO is costly, opaque; HyperSkill evolves a hypergraph-indexed skill library.

Task Definition

Let $\pi\theta$ be a **frozen omnimodal backbone** (θ is never updated) and let train / test sets hold triples (q, a^*, m): a query, its gold answer and a modality tag $m \in$ {text, image, pdf, video, audio}.

We seek a **weighted skill hypergraph** $H = (V, E, w)$: every node carries a name, a description and an **advantage** $u_i \in [-1, 1]$; every hyperedge groups ≥ 2 nodes into a joint composition with weight $w(e_j)$.

We also seek a **retriever** $R(q, m, H) \rightarrow S_q \subset V$, while a **hypergraph update U** replaces the gradient step in every round t:

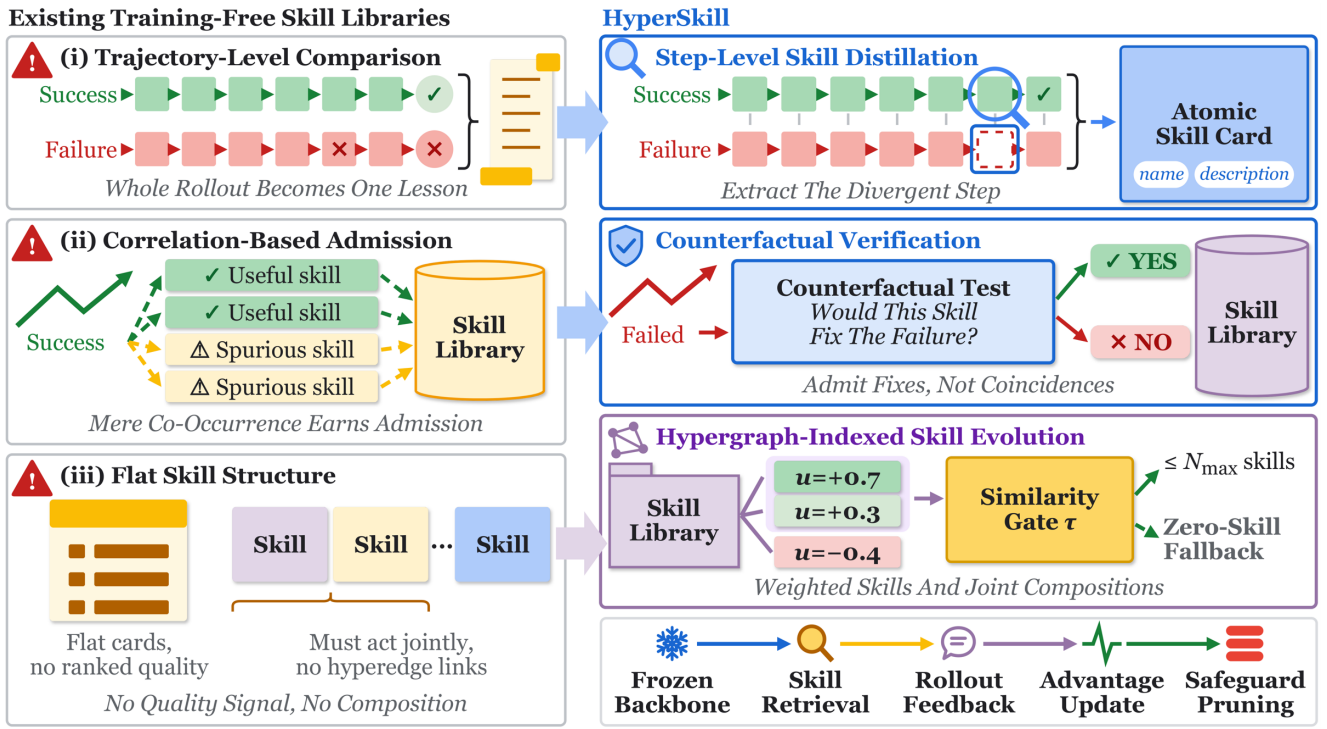


Figure 2: Failure modes (i)–(iii) of existing training-free skill libraries vs. HyperSkill's remedies.

$$\left(\mathcal{H}^{\star},\, \mathcal{R}^{\star}\right) \;=\; \underset{\mathcal{H},\,\mathcal{R}}{\operatorname{argmax}}\; \mathbb{E}_{(q,a^*,m)\sim\mathcal{D}}\Big[\, r\big(\,\pi_\theta(\cdot\,|\,q, m, \mathcal{R}(q, m, \mathcal{H}))\,;\, a^*\,\big)\,\Big]\,, \quad \mathcal{H}_{t+1} = \mathcal{U}(\mathcal{H}_t; \mathcal{B}_t)\,, \tag{1}$$

r is the token-level F1 reward. Unlike GRPO, the optimisation variable is **(H, R) rather than θ**: gradient updates become an iterative, auditable rewrite of the library.

Three Main Challenges

Existing training-free skill libraries face three key challenges:

i **Trajectory-level comparison**
Existing libraries compare at the trajectory level, so every divergent step is fused into one lesson ("the whole rollout becomes one lesson").

ii **Correlation-based admission**
Candidates are admitted by co-occurrence with successful rollouts, so the library keeps patterns that track reward but do not fix the failure.

iii **Flat skill structure**
Skills are stored as dictionaries or uniform-quality lists, which cannot express that several skills must act jointly or give a per-skill quality signal.

2 PART

Contributions

NEURAL INFORMATION PROCESSING SYSTEMS

Contributions

1

Training-Free Omnimodal GRPO

We propose **HyperSkill**, which keeps the backbone frozen and replaces gradient updates with auditable edits to a Hypergraph-Indexed Skill-Library: skills are advantage-bearing nodes, compositions are hyperedges.

2

Counterfactual distillation + gated retrieval

C1 admits step-level skills only after a counterfactual check and learns a protocol skill; C2 ranks hyperedge-linked skills by advantage behind a gate with zero-skill fallback; C3 makes the backbone commit to an answer.

3

+13.0 average F1

On 11 **text / VL / omnimodal benchmarks**, HyperSkill is best or tied on every dataset, lifting average F1 from 29.0 (XSkill) to 42.0, with libraries costing at most $0.88 to train.

3

PART

Methodology

NEURAL INFORMATION PROCESSING SYSTEMS

Method: HyperSkill

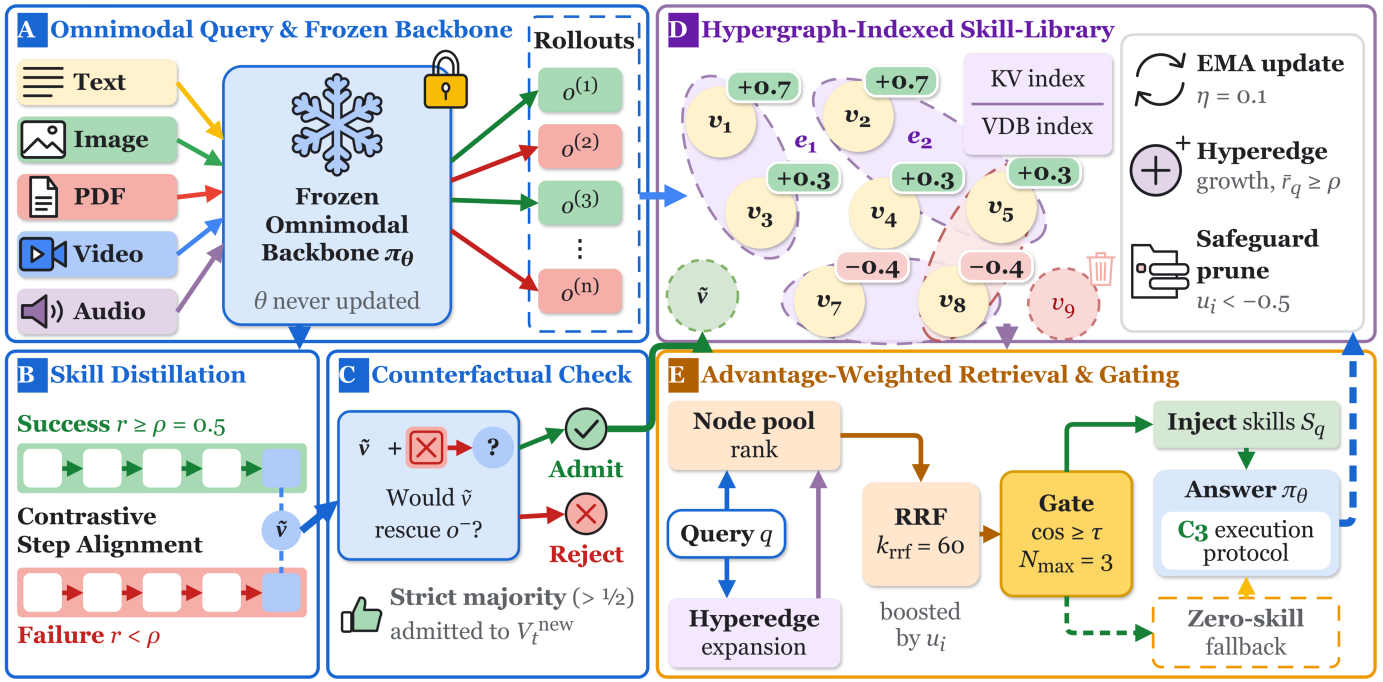


Figure 3: Methodology of HyperSkill: C1 builds (B–C), C2 evolves & retrieves (D–E) the skill hypergraph; C3 governs execution.

C1
$$\widetilde{\mathcal{V}}_q = \Phi\Big(\{ \Psi(o^{(\ell)}) : r(o^{(\ell)}; a^*) \geq \rho \}_{\ell=1}^{n}, \ \{ \Psi(o^{(\ell)}) : r(o^{(\ell)}; a^*) < \rho \}_{\ell=1}^{n} \Big), \qquad (2)$$

C1
$$\mathcal{V}_t^{\text{new}} = \Big\{ \tilde{v} \in \widetilde{\mathcal{V}}_q \ : \ \frac{1}{|\mathcal{P}_q|} \sum_{o^- \in \mathcal{P}_q} \mathbf{1}\big[\pi_\theta(\cdot \mid \tilde{v}, q, a^*, o^-) = \text{YES} \big] \ > \ \tfrac{1}{2} \Big\}, \qquad |\mathcal{P}_q| \leq K = 2. \quad (3)$$

C2
$$\hat{A}_q = \tanh\Big(\tfrac{\bar{r}_q - \bar{r}_{\mathcal{B}_t}}{\sigma_{r,\mathcal{B}_t} + \epsilon} \Big), \quad u_i \leftarrow \text{clip}_{[-1,1]}\big((1-\eta)\, u_i + \eta\, \hat{A}_q \big) \ \forall v_i \in \mathcal{S}_q, \qquad (4)$$

C2
$$\text{score}(v_i; q) = \frac{1 + \frac{1}{2}(u_i + 1)}{k_{\text{rrf}} + \text{rank}_f(v_i; q)}, \qquad \mathcal{S}_q^{\text{pool}} = \text{Top-}k_{\cos}(\mathcal{V}) \cup \text{expand}_q(\mathcal{E}), \qquad (6)$$

C2
$$\mathcal{S}_q = \text{Top-}N_{\max}\Big(\text{score}(\cdot\,; q), \ \{ v_i \in \mathcal{S}_q^{\text{pool}} \ : \ \langle f(q), f(v_i) \rangle \geq \tau \} \Big), \qquad N_{\max} = 3, \qquad (7)$$

Optimise the library, not θ

The backbone stays frozen; an auditable library edit ΔH replaces the GRPO gradient step Δθ.

C1 Training-Time Skill Distillation

Contrastive step alignment (Eq. 2) harvests **divergent step-level skills** from success vs. failure rollouts; a **counterfactual probe** (Eq. 3) admits a skill only on a strict majority of YES votes; a protocol skill is learned from failures.

C2 Hypergraph Evolution & Gated Retrieval

Rollout rewards update per-node **advantages** by EMA (Eq. 4); joint skills grow **hyperedges**; low-advantage nodes are pruned. At test time **advantage-weighted RRF** (Eq. 6) + a similarity gate τ, cap $N_{max} = 3$ and a **zero-skill fallback** (Eq. 7) decide what is injected.

C3 Execution Protocol

Never edits H; per-dataset options (tool prompts, lenient parser, final-answer turn, evidence & voting) make the frozen backbone **commit to an answer**.

4

PART

Experimental Results

NEURAL INFORMATION PROCESSING SYSTEMS

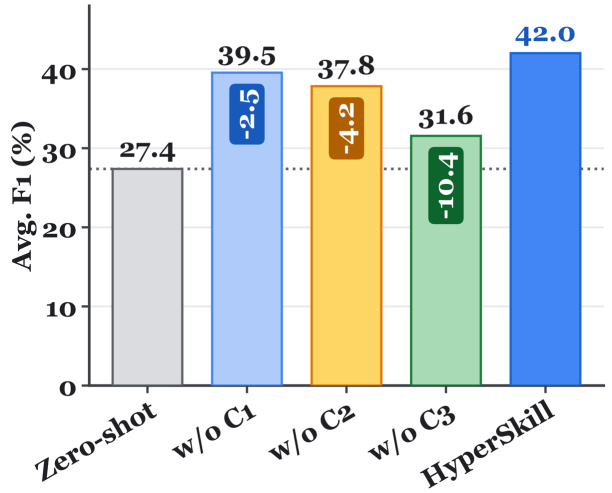
Main Results

Modality	Dataset	Zero-shot	T.F.GRPO	SkillCraft	AutoSkill	EvoSkill	M.-Skills	XSkill	HyperSkill
Text	AIME2024	10.0	10.0$_{\pm 0.0}$	13.3$_{+3.3}$	13.3$_{+3.3}$	10.0$_{\pm 0.0}$	13.3$_{+3.3}$	6.7$_{-3.3}$	**40.0**$_{+30.0}$
	AIME2025	33.3	26.7$_{-6.6}$	30.0$_{-3.3}$	33.3$_{\pm 0.0}$	33.3$_{\pm 0.0}$	26.7$_{-6.6}$	30.0$_{-3.3}$	**53.3**$_{+20.0}$
	WebWalkerQA	20.7	16.9$_{-3.8}$	16.5$_{-4.2}$	20.7$_{\pm 0.0}$	18.6$_{-2.1}$	16.3$_{-4.4}$	14.8$_{-5.9}$	**23.3**$_{+2.6}$
VL-Web	FVQA	21.1	27.0$_{+5.9}$	17.8$_{-3.3}$	31.3$_{+10.2}$	23.9$_{+2.8}$	19.5$_{-1.6}$	26.6$_{+5.5}$	**41.7**$_{+20.6}$
	InfoSeek	22.8	27.4$_{+4.6}$	15.2$_{-7.6}$	22.3$_{-0.5}$	16.8$_{-6.0}$	22.7$_{-0.1}$	33.4$_{+10.6}$	**44.8**$_{+22.0}$
	SimpleVQA	26.5	19.3$_{-7.2}$	19.8$_{-6.7}$	27.7$_{+1.2}$	33.7$_{+7.2}$	28.6$_{+2.1}$	33.8$_{+7.3}$	**37.4**$_{+10.9}$
	LiveVQA	55.0	56.3$_{+1.3}$	59.9$_{+4.9}$	64.4$_{+9.4}$	59.4$_{+4.4}$	58.5$_{+3.5}$	59.6$_{+4.6}$	**65.7**$_{+10.7}$
VL-Doc	DocBench	32.8	26.1$_{-6.7}$	16.0$_{-16.8}$	32.9$_{+0.1}$	32.8$_{\pm 0.0}$	35.1$_{+2.3}$	33.1$_{+0.3}$	**39.2**$_{+6.4}$
	MMLongB.	19.4	22.6$_{+3.2}$	9.8$_{-9.6}$	12.4$_{-7.0}$	15.4$_{-4.0}$	10.1$_{-9.3}$	23.0$_{+3.6}$	**32.1**$_{+12.7}$
Omni	OmniVideoB.	16.0	6.4$_{-9.6}$	14.9$_{-1.1}$	12.2$_{-3.8}$	16.0$_{\pm 0.0}$	5.1$_{-10.9}$	17.2$_{+1.2}$	**31.5**$_{+15.5}$
	WorldSense	43.8	47.5$_{+3.7}$	43.8$_{\pm 0.0}$	43.8$_{\pm 0.0}$	43.8$_{\pm 0.0}$	49.1$_{+5.3}$	40.6$_{-3.2}$	**53.1**$_{+9.3}$
	Avg.	27.4	26.0$_{-1.4}$	23.4$_{-4.0}$	28.6$_{+1.2}$	27.6$_{+0.2}$	25.9$_{-1.5}$	29.0$_{+1.6}$	**42.0**$_{+14.6}$

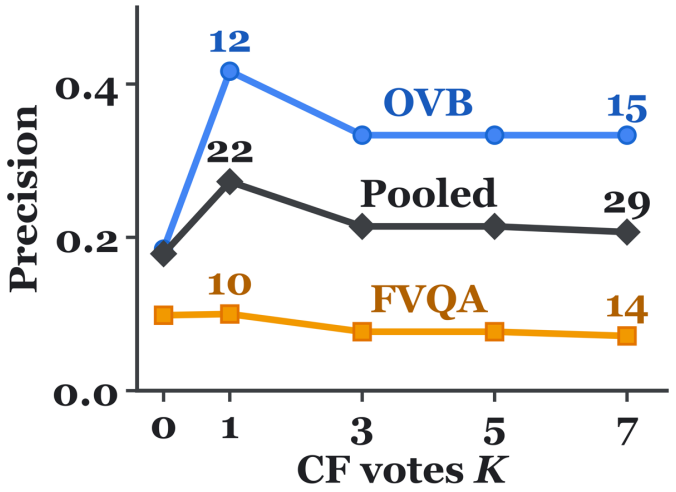
HyperSkill-Bench: 11 public datasets in 4 regimes (Text, VL-Web, VL-Doc, Omni), 636 test questions, token-level F1. All methods share a frozen **gemini-2.5-flash-lite** backbone, the same tools and scoring; six training-free skill baselines use the same 128 training questions.

Best or tied on all 11 datasets; average F1 29.0 (XSkill) → **42.0**: **+13.0 abs. / +44.9% rel.**, +14.6 over Zero-shot. Largest margins AIME2024 +26.7, AIME2025 +20.0, OmniVideoBench +14.3; regime leads Text +16.4, Omni +9.2, VL-Web +6.7, VL-Doc +6.6.

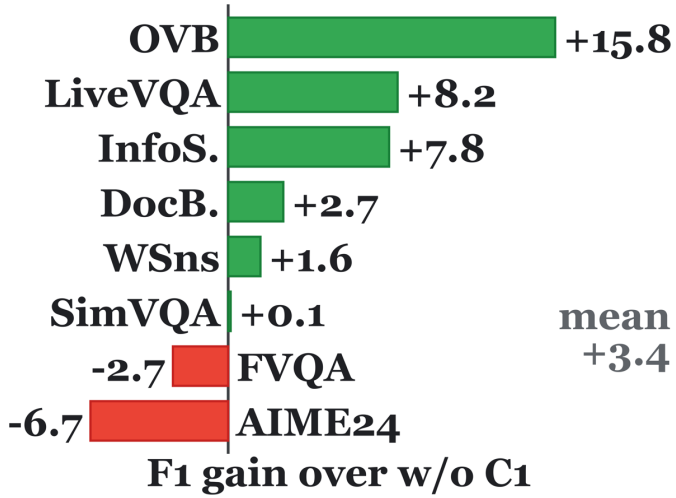
Ablation & Analysis (RQ2–RQ5)



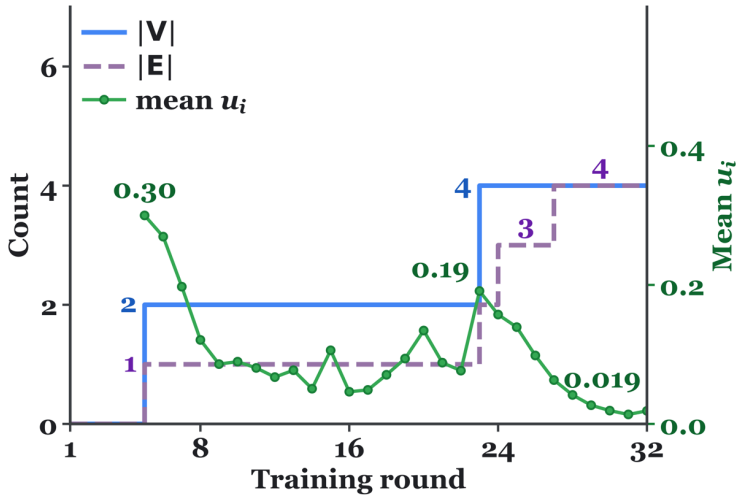
(a) Ablation: 42.0 → 39.5 / 37.8 / 31.6 (w/o C1 / C2 / C3)



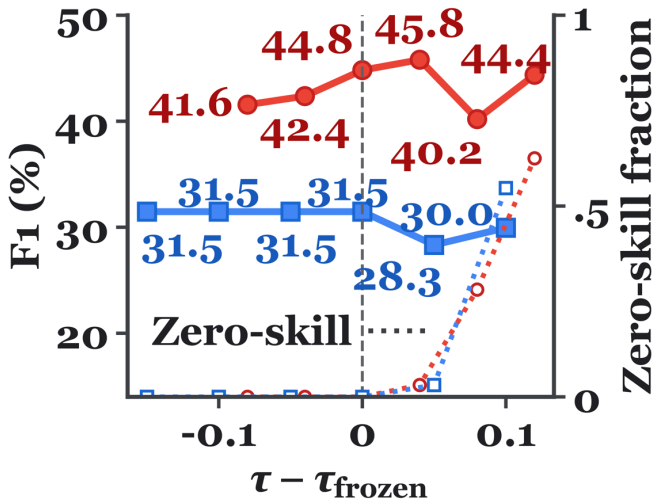
(b) Probe precision vs. K (labels: #accepted; AIME24 none)



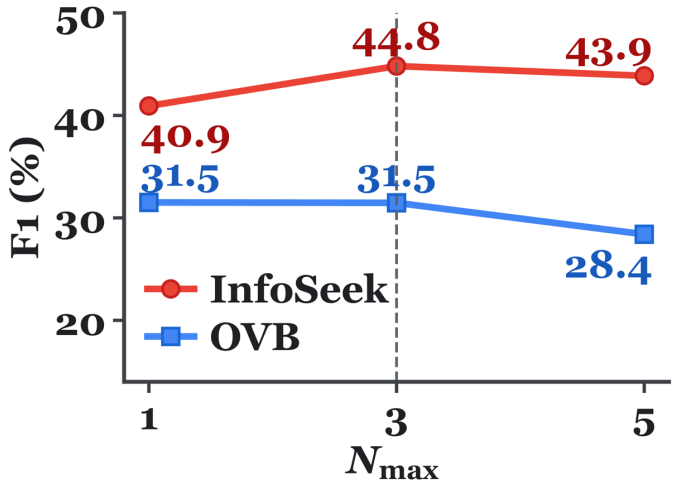
(c) F1 gain over w/o C1 on the 8 differing datasets (mean +3.4)



(d) Library size and mean u_i over 32 training rounds

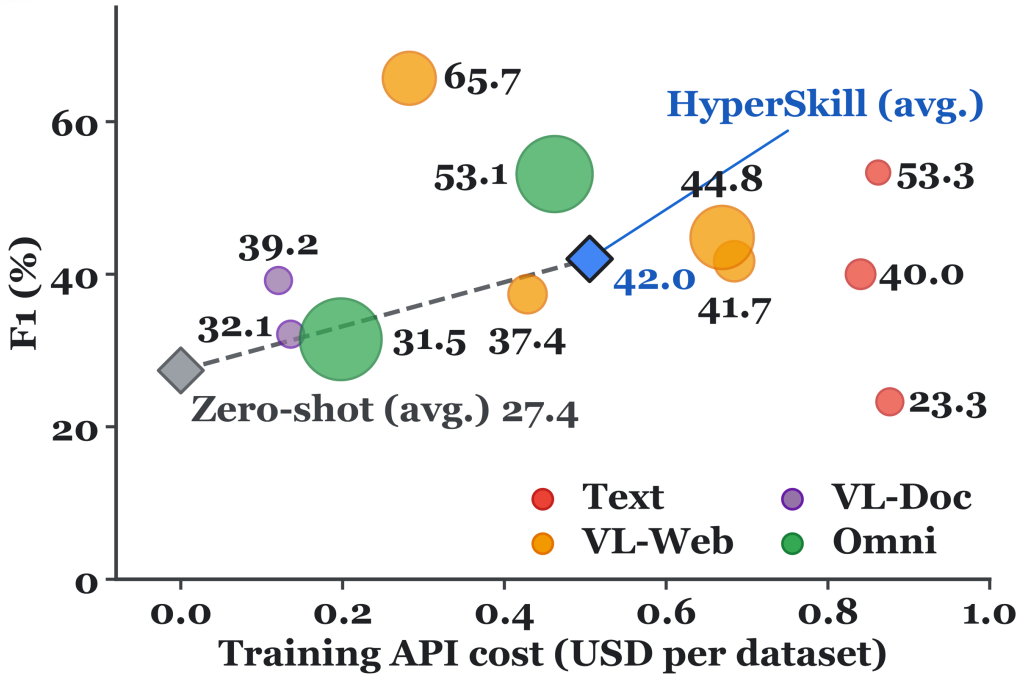


(e) F1 & zero-skill fraction vs. offset of τ from its frozen value

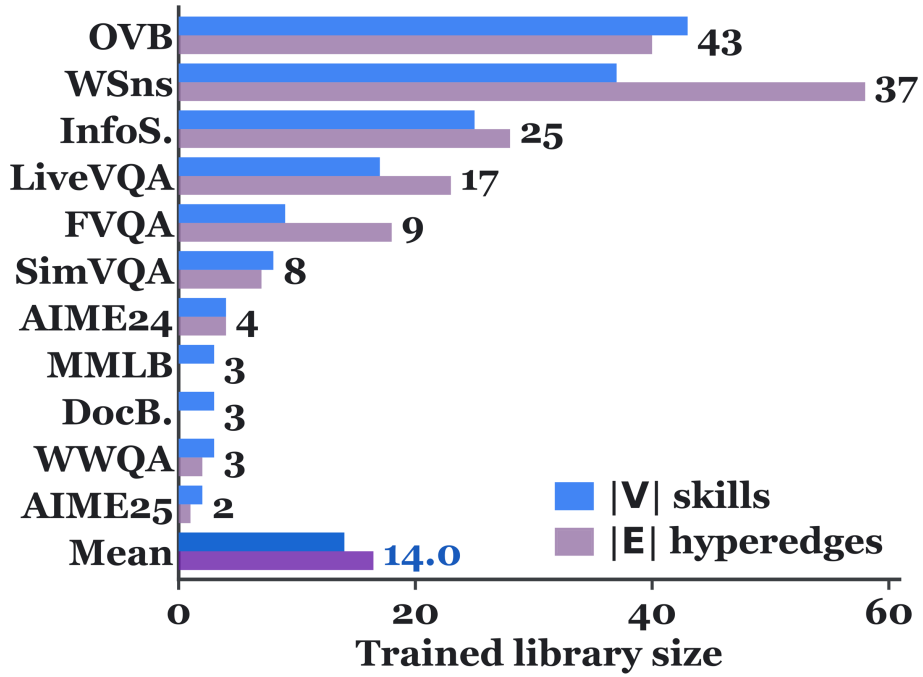


(f) F1 vs. capacity cap N_max (InfoSeek, OVB; frozen at 3)

Cost & Efficiency (RQ6)



(a) Training API cost vs. F1 per dataset, coloured by regime (bubble ∝ |V|)



(b) Library size per dataset: skills |V| (dark), hyperedges |E| (light)

Counterfactual distillation keeps a **compact library**: 154 skills vs. 407 without the probe (38%); one CF vote raises pooled probe precision from 0.179 to 0.273.

Libraries hold 2–43 skills (mean 14.0) and 0–58 hyperedges; the frozen gate settings lie in a stable region.

$0.12–$0.88 training cost per dataset (mean $0.51, total $5.56)

$2.34 one full test pass over all 11 datasets

≤ ~3 s added latency vs. Zero-shot

40th Conference on Neural Information Processing Systems

Sydney (Main)
Dec 6 – Dec 12, 2026

NeurIPS 2026 Poster

Thanks for Listening

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Code & Data https://github.com/LHRLAB/HyperSkill

Contact haoran.luo@ntu.edu.sg

VINUNIVERSITY