Anti-Collapse Protocol: Synthetic Recursion Without Collapse via Entropy Banding, Tail Preservation, Isotropy, and Red-Queen Critics

Anonymous ACL submission

Abstract

Premise. High-quality human text is becoming a finite bottleneck for scaling language-model training, pushing the field toward **synthetic data**. Yet naïve recursive self-feeding induces **model collapse**: rare tokens disappear, outputs converge to stylistic monoculture, calibration drifts, and rare knowledge decays first.

We propose the **Anti-Collapse Protocol (ACP)**, a **provenance-agnostic** recipe for repeated synthetic fine-tuning that preserves ***heterozygosity*** without detecting "AI vs. human." ACP interlocks four information-geometry gates: **(T1) Entropy Banding** admits mid-difficulty, learnable batches; **(T2) Tail-Mass & Novelty Floors** with near-duplicate caps preserve rare-token throughput and out-of-support n-grams; **(T3) Domain Isotropy** with temporal stratification prevents style/recency lock-in; and **(T4) Red-Queen Critics** over a mechanism-diverse teacher lattice preserve multiple correct routes while penalizing sameness. ACP runs as a closed loop: **Synthesize** $\rightarrow$ **Screen** $\rightarrow$ **Train** $\rightarrow$ **Measure** $\rightarrow$ **Adapt**, with pre-registered tripwires.

We evaluate ACP in an academic repeated fine-tuning regime using a **Llama-3 8B + LoRA** student and teachers **Llama-3 70B**, **Mixtral** 8×22**B**, and **Qwen2-72B**. Across four synthetic windows **R1**$\rightarrow$**R4**, ACP maintains entropy bands, tail mass, novelty, isotropy, and route diversity with **zero tripwire breaches**; the Recursive-Self baseline fails by **R2–R3** on duplicate and calibration thresholds. GLUE macro shows no degradation **(+0.3%)**, and ablations reproduce the predicted failures: degenerate batches, tail death, style bubbles, and route monoculture, **matching the targeted collapse mechanisms gate by gate under controlled recursion**. Code and Demo.

The Anti-Collapse Protocol

- ***Information-geometry framing:*** Collapse is an entropy/tail/novelty/style failure—not a provenance failure. ACP regulates entropy bands, tail mass, novelty rate, isotropy, and mechanism plurality rather than "AI vs. human" origin.
- ***ACP core—four constraints:*** A closed loop: **Synthesize** $\rightarrow$ **Screen** (T1–T4) $\rightarrow$ **Train** $\rightarrow$ **Measure** $\rightarrow$ **Adapt**. T1 Entropy Banding, T2 Tail–Mass & Novelty Floors, T3 Domain Isotropy, and T4 Red–Queen Critics jointly preserve heterozygosity.
- ***Compute-feasible instantiation:*** ACP uses a **Llama-3-8B** LoRA/QLoRA student with a teacher lattice: **L3-70B**, **Mixtral 8×22B**, and **Qwen2-72B**, across four synthetic windows R1$\rightarrow$R4.
- ***Pre-registered diagnostics & tripwires:*** We track tail health (Zipf-tail mass, Heaps b, unseen n-grams), diversity (style entropy, repetition/burstiness), coverage (domain/time KL, W_1), calibration (ECE/Brier), and route health (entropy, MI).
- ***Mechanism diversity:*** Direct, CoT, plan-first, retrieve-first, and symbolic/pseudocode routes are maintained via route-aware mixing, counterfactual recomposition, and rotating Red–Queen critics that penalize sameness.
- ***Empirical stability under recursion:*** Across R1–R4, ACP preserves tail mass/novelty, reduces repetition, bounds domain/time drift, sustains route entropy, and improves CVaR–NLL/calibration, while *Recursive-Self* shows collapse signatures.
- ***Causal attribution via ablations:*** No-T1 $\Rightarrow$ trivial/degenerate batches; No-T2 $\Rightarrow$ tail shrinkage & clone inflation; No-T3 $\Rightarrow$ style/recency bubbles; No-T4 $\Rightarrow$ route monoculture.
- ***Reviewer-auditable protocol:*** ACP provides transparent knobs, pre-registered thresholds, and release-ready figures/logs for growing synthetic data while preserving rare knowledge, plural styles, and diverse mechanisms.

1 The Anti-Collapse Protocol: Necessity & Urgency

Data–exhaustion premise. Multiple independent analyses suggest that the ***quality-adjusted stock of public, human-authored web text*** is finite and may be substantially consumed within the next few frontier training cycles. Forecasts combining historical dataset growth with compute-scaled training trends estimate an accessible stock of roughly ~ 4–5×10^{14} tokens and project that frontier-scale runs may approach or exceed this stock by the late-2020s to early-2030s [Villalobos et al., 2024][1]. In parallel, compute-optimal scaling studies show

[1] https://epochai.org/blog/data-scaling-and-when-we-run-out

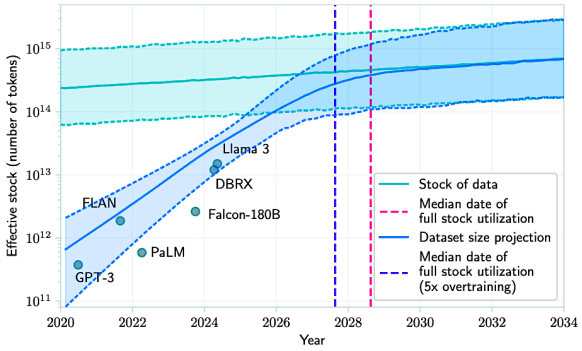


Figure 1: **Forecasted exhaustion of public human text for LLM training. Teal:** estimated effective stock of quality-adjusted, de-duplicated public human text on the web, approximately $4-5 \times 10^{14}$ tokens. **Blue:** projected dataset sizes of frontier training runs extrapolated from historical scaling trends. Vertical markers indicate median crossover dates under conservative and aggressive over-training assumptions. **Key takeaway:** projected training demand intersects estimated human-text supply in the late-2020s/early-2030s, motivating controlled synthetic-data regimes. Adapted from [Villalobos et al., 2024; AI, 2023].

that capability gains historically depended on increasing both parameter count and training tokens, structurally increasing demand for fresh data as scale grows [Kaplan et al., 2020; Hoffmann et al., 2022]. This creates a ***methodological shift***: future progress increasingly depends on procedures that can **(i) generate new signal**, **(ii) curate it for quality and diversity**, and **(iii) stabilize learning under repeated reuse**. Fig. 1 visualizes this crossover.

Synthetic recursion is increasingly unavoidable. As high-quality human text plateaus, synthetic corpora are already used to fill domain gaps, satisfy privacy constraints, augment instruction tuning, and extend scaling regimes. At the same time, future crawls are increasingly likely to contain model-generated content, creating ***recursive training loops*** in which models partially learn from earlier model emissions. External audits already document rapid proliferation of AI-generated web content, increasing the probability of inadvertent self-consumption during future crawling. Synthetic recursion is therefore ***not only a design choice, but an emerging property of the data ecosystem***.

Collapse is a distributional failure. Naïve recursive training induces ***model collapse***: rare modes disappear, stylistic diversity contracts, repetition increases, and calibration drifts [Shumailov et al., 2024a]. These failures usually appear first in the ***tails***: rare entities, unusual syntax, minority languages, low-frequency reasoning paths, and atypical knowledge distributions. Later, the model drifts toward head-distribution clichés and stylistic monoculture. Thus, collapse is not only semantic; it is ***distributional***, involving **(i) entropy**, **(ii) support coverage**, **(iii) novelty**, **(iv) calibration**, and **(v) mechanism diversity**.

Why provenance is the wrong control knob. In mixed ecosystems, "AI vs. human" labels are noisy, adversarially gameable, and often orthogonal to information quality. A synthetic paragraph may contain useful novelty; a human paragraph may be duplicated boilerplate. Recursive stability therefore cannot be guaranteed by provenance filtering alone. The more relevant question is whether the training distribution preserves the ***statistical conditions under which learning remains diverse, calibrated, and information-rich***. This framing aligns with broader concerns that representation quality and reuse dynamics matter more than simple origin labels [Bender et al., 2021].

Table 1: **Prior remedies grouped into four major directions.** Existing work addresses important pieces of the recursion problem, but largely in isolation.

Hygiene & Mixture Quality	**Clean & decontaminate.** Deduplication, eval decontamination, per-source reweighting, mixture rebalance, learned/rule-based quality scoring, and allow/deny lists. Anchors: **C4** [Raffel et al., 2020], **The Pile** [Gao et al., 2021], **RefinedWeb** [Penedo et al., 2023], **DOLMA** [Soldaini et al., 2024], dedup/privacy [Kandpal et al., 2022], contamination auditing [Deng et al., 2024a], **FineWeb** [Penedo et al., 2023], **RedPajama** [Weber et al., 2024].
Targeted Acquisition Loops	**Mine what is missing.** Uncertainty/disagreement sampling, curriculum/self-paced learning, hard-slice targeting, adversarial data collection, and prompt evolution. Anchors: active learning [Settles, 2009], curriculum learning [Bengio et al., 2009], self-paced learning [Kumar et al., 2010], **Self-Instruct** [Wang et al., 2023], **Evol-Instruct/WizardLM** [Xu et al., 2023], **Dynabench**, **ANLI** [Nie et al., 2020].
Anti-Collapse Objectives & Decoding	**Fight mode collapse.** Unlikelihood training, diversity-oriented decoding, novelty/tail sampling, contrastive search, entropy regularization, and explicit repetition penalties. Anchors: degeneration/nucleus sampling [Holtzman et al., 2020], unlikelihood [Welleck et al., 2020], typical decoding [Meister et al., 2023], diverse beam search [Vijayakumar et al., 2018], contrastive search [Su et al., 2022].
Theory, Metrics & Tripwires	**Define & detect.** Collapse theory, tail/support diagnostics, Zipf/Heaps statistics, calibration metrics, scaling laws, and pre-registered halting criteria. Anchors: recursive collapse [Shumailov et al., 2024a], collapse analyses, scaling laws [Kaplan et al., 2020], Chinchilla [Hoffmann et al., 2022], data-limit concerns [Bender et al., 2021], Zipf/Heaps [Zipf, 1949; Heaps, 1978], calibration [Guo et al., 2017].

Prior remedies are useful but fragmented. Table 1 organizes prior approaches into four broad directions: **data hygiene**, **targeted acquisition**, **anti-collapse objectives**, and **theory/diagnostic frameworks**. Each direction addresses part of the recursion problem: hygiene improves corpus quality; acquisition methods mine missing or adversarial regions; anti-collapse objectives reduce repetition and degeneration; and theory/metric work explains how collapse can be detected. However, these strands are usually studied independently. What remains missing is a unified protocol that jointly controls **(i) which synthetic samples en-**

ter training, (ii) whether tail and novelty statistics remain healthy, and **(iii) when the pipeline should adapt before recursive drift compounds**.

Need for a closed-loop anti-collapse protocol. The missing object is not another detector, filter, or decoding heuristic, but a ***closed-loop protocol*** that treats repeated synthetic fine-tuning as a controlled data-generation process. Such a protocol should **screen incoming synthetic data**, **monitor early indicators of collapse**, and **adapt generation or filtering** before instability compounds across rounds. The **Anti-Collapse Protocol (ACP)** is our instantiation of this idea. Rather than filtering by provenance, ACP regulates the ***statistical structure of the training distribution itself***.

Scope and positioning. We do not claim to solve frontier-scale synthetic pretraining. Instead, we study a controlled repeated fine-tuning regime where collapse mechanisms can be measured and intervened on directly. Our claim is narrower: recursive synthetic fine-tuning becomes more stable when governed through ***explicit information-geometric constraints*** rather than provenance filtering alone. The next section formalizes ACP, including its gates, diagnostics, and recursive control mechanisms.

2 The Anti-Collapse Protocol (ACP)

Overview. ACP treats recursive synthetic fine-tuning as a ***closed-loop data-admission and distribution-control problem***. At recursive window w, a generator produces a candidate pool $\mathcal{C}_w$, from which ACP selects an accepted subset $\mathcal{A}_w \subseteq \mathcal{C}_w$ for training. The key design question is not whether a sample is *human* or *synthetic*, but whether it preserves the statistical conditions under which recursive learning remains stable: **learnability**, **tail support**, **coverage balance**, and **mechanism diversity**. This reframes recursive collapse as an ***information-geometry failure*** rather than a provenance failure [Shumailov et al., 2024a; Deng et al., 2024b].

Formally, ACP performs constrained recursive data admission:

$$\mathcal{A}_w^\star = \arg\max_{\mathcal{A}\subseteq\mathcal{C}_w} \mathcal{U}(\mathcal{A};\mathcal{S}_{w-1}) \quad \text{s.t.} \quad \mathbf{g}(\mathcal{A}) \succeq \mathbf{0},$$

where

$$\mathbf{g}(\mathcal{A}) = \begin{bmatrix} g_{\text{ent}}(\mathcal{A}) & : & \text{learnability} \\ g_{\text{tail}}(\mathcal{A}) & : & \text{tail/novelty support} \\ g_{\text{iso}}(\mathcal{A}) & : & \text{domain–time coverage} \\ g_{\text{route}}(\mathcal{A}) & : & \text{mechanism diversity} \end{bmatrix}.$$

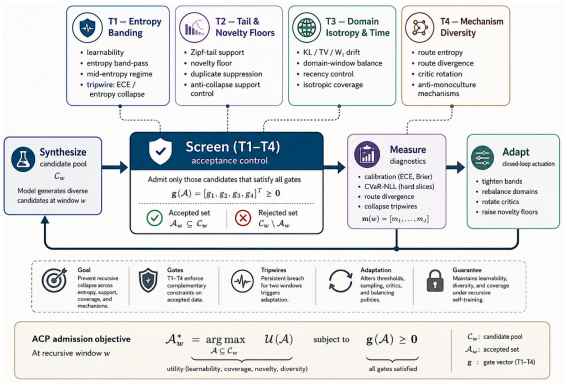


Figure 2: **ACP as closed-loop recursive control.** At window w, synthetic candidates are generated, screened through four gates, used for training, measured through diagnostics, and reweighted before the next window. **T1** controls learnability, **T2** preserves tail/novelty support, **T3** enforces domain–time coverage, and **T4** maintains mechanism diversity. ACP regulates the geometry of information rather than provenance.

The student then updates recursively:

$$\mathcal{S}_w = \text{Train}(\mathcal{S}_{w-1}, \mathcal{A}_w^\star).$$

ACP therefore acts as a ***recursive stabilizer***: it regulates the geometry of admitted information before collapse signatures compound across windows. The protocol operates as

$$\textsc{Synthesize} \rightarrow \textsc{Screen} \rightarrow \textsc{Train} \rightarrow \textsc{Measure} \rightarrow \textsc{Adapt},$$

where SCREEN applies four interoperable gates: **(T1)** Entropy Banding, **(T2)** Tail-Mass & Novelty Floors, **(T3)** Domain Isotropy with Temporal Stratification, and **(T4)** Red-Queen Critics.

2.1 T1: Entropy Banding

Motivation. Recursive pipelines naturally overproduce two pathological regimes: **(i)** low-entropy rote text that reinforces dominant templates and memorized continuations, and **(ii)** high-entropy synthetic noise that injects unstable gradients without reliable support expansion. ACP therefore restricts admission to ***mid-entropy, learnable regions*** where uncertainty remains informative but structurally coherent.

For sequence $x = (t_1, \ldots, t_{|x|})$, define sequence entropy

$$H_{\text{seq}}(x) = -\frac{1}{|x|}\sum_{i=1}^{|x|} \log p(t_i \mid x_{<i}).$$

Given a held-out probe set $\mathcal{P}_w$, ACP defines a window-specific entropy band by probe quantiles:

$$\underline{h}_w = Q_\alpha(H_{\text{seq}};\mathcal{P}_w), \qquad \overline{h}_w = Q_{1-\alpha}(H_{\text{seq}};\mathcal{P}_w),$$

$$\mathcal{B}_w = [\underline{h}_w, \overline{h}_w].$$

A batch $B \subseteq \mathcal{C}_w$ passes T1 only if its band-pass rate exceeds the admission threshold:

$$g_{\text{ent}}(B) = \Pr_{x \in B}\left[H_{\text{seq}}(x) \in \mathcal{B}_w\right] - \theta_H \geq 0.$$

The T1-screened pool is therefore

$$\mathcal{A}_w^{(1)} = \{B \subseteq \mathcal{C}_w : \; g_{\text{ent}}(B) \geq 0\}.$$

Equivalently, T1 acts as a ***band-pass filter on useful uncertainty***:

$$\underbrace{\text{rote repetition}}_{\text{low entropy}} \longrightarrow \underbrace{\text{learnable uncertainty}}_{\text{mid entropy}} \longrightarrow \underbrace{\text{unstable noise}}_{\text{high entropy}}.$$

Very predictable sequences contribute little corrective signal, while extremely uncertain sequences often behave like noisy supervision. The middle regime maximizes useful gradient density: it is uncertain enough to teach, but structured enough to learn from. This connects ACP to uncertainty-based acquisition, curriculum learning, self-paced optimization, and the ***Minimum Description Length (MDL)*** principle: useful supervision should remain compressible enough to learn from, but not so predictable that it contributes no new information [Settles, 2009; Bengio et al., 2009; Kumar et al., 2010; Rissanen, 1978, 1989]. Fig. 3 visualizes the entropy band enforced by T1, separating rote repetition and unstable noise from the learnable mid-entropy regime.

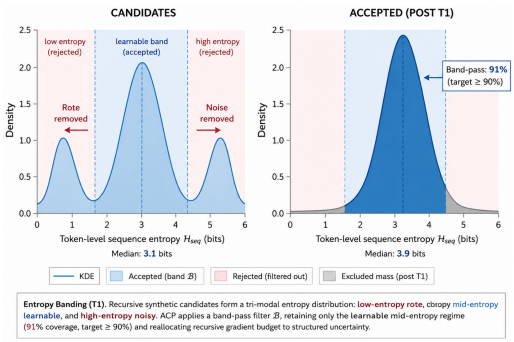


Figure 3: **Entropy Banding (T1).** Candidate synthetic data split into three regimes: **(left)** low-entropy rote repetition dominated by memorized templates, **(middle)** structured mid-entropy samples with useful corrective uncertainty, and **(right)** unstable high-entropy noise. ACP keeps only the learnable middle band, reallocating recursive gradient budget toward informative supervision rather than repetition or incoherence.

2.2 T2: Tail-Mass & Novelty Floors

Motivation. Recursive collapse appears first in the tails. Rare entities, unusual syntax, minority styles, and low-frequency reasoning patterns disappear long before headline accuracy degrades. T2 therefore preserves ***effective support coverage*** through three coupled constraints: **tail throughput**, **out-of-support novelty**, and **duplicate suppression**.

Let $\mathcal{T}_q$ denote the least-frequent q-fraction of vocabulary types under Zipf ranking:

$$\mathcal{T}_q = \{t : r(t) > \lfloor (1-q)|\mathcal{V}| \rfloor\}.$$

For batch B, define

$$\text{TailMass}_q(B) = \sum_{t \in \mathcal{T}_q} p_B(t), \qquad \text{Novelty}_n(B) = \frac{|U_n(B)|}{|\text{ngrams}_n(B)|},$$

where $U_n(B)$ contains previously unseen n-grams. ACP jointly enforces

$$g_{\text{tail}}(B) = \min\Big\{\text{TailMass}_q - \eta_q, \; \text{Novelty}_n - \gamma_n, \; \kappa - \text{Dup}(B)\Big\} \geq 0.$$

where $\text{Dup}(B)$ is the maximum near-duplicate rate under lexical, semantic, or template matching.

The three terms regulate complementary failure modes:

$$\begin{aligned} \text{tail floors} &\rightarrow \text{rare support} \\ \text{novelty floors} &\rightarrow \text{support expansion} \\ \text{duplicate caps} &\rightarrow \text{clone suppression} \end{aligned}$$

Together, T2 maintains ***support heterozygosity*** under recursion—by analogy to population diversity, where heterozygosity measures the preservation of multiple allelic variants rather than clonal dominance [Nei, 1973; Hartl and Clark, 1997]. Tail floors ensure that rare regions continue receiving gradient budget; novelty floors inject genuinely new compositional structure; duplicate caps prevent synthetic loops from inflating apparent diversity through paraphrastic clones. This directly targets the support-collapse dynamics predicted by recursive-collapse theory [Shumailov et al., 2024a], connects to deduplication and contamination-control concerns [Kandpal et al., 2022; Deng et al., 2024a], and aligns with Zipf/Heaps diagnostics of vocabulary health [Zipf, 1949; Heaps, 1978]. Fig. 4 illustrates how T2 preserves Zipf-tail support while capping duplicate amplification under recursive reuse.

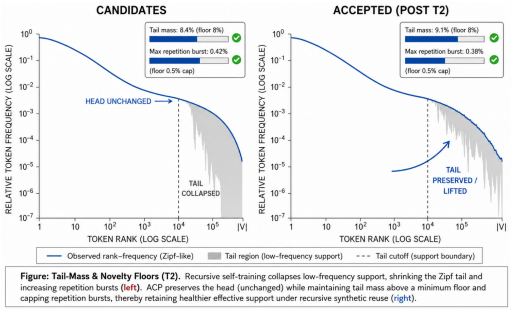


Figure 4: ***Tail-Mass & Novelty Floors (T2). Recursive self-training preserves the head but progressively erodes the low-frequency tail***: rare tokens disappear, novelty growth slows, and near-duplicate templates expand. ***T2 counteracts this collapse*** by enforcing a minimum Zipf-tail mass, requiring genuinely unseen n-grams, and capping duplicate bursts. The accepted distribution therefore preserves ***healthier effective support under repeated synthetic reuse***.

2.3 T3: Domain Isotropy with Temporal Stratification

Motivation. Even when recursive samples remain learnable and novel, training can still collapse toward dominant domains, formats, or recent synthesis windows. Recursive systems naturally drift toward ***style monoculture*** and ***recency lock-in***. T3 therefore regulates ***coverage geometry*** across domains and recursive windows.

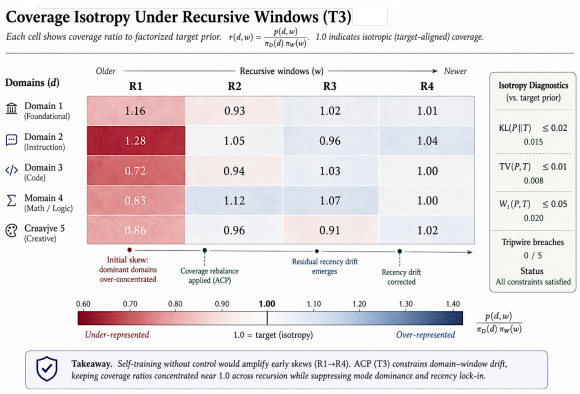


Figure 5: ***Domain Isotropy with Temporal Stratification (T3). ACP maintains balanced coverage across domains and recursive windows***, preventing gradient concentration onto dominant styles or recent generations. Cells near unity indicate ***stable alignment with target priors***, while concentrated regions correspond to ***style monoculture and recency lock-in***.

Let $\hat{p}_w(d)$ and $\hat{p}_w(r)$ denote the accepted domain and recursive-window distributions at synthesis window w, with target priors π_D and π_R. To prevent marginal balance from hiding cell-level collapse, we also track the joint domain–window distribution $\hat{p}_w(d, r)$. ACP constrains

$$\mathrm{KL}(\hat{p}_w(d) \,\|\, \pi_D) \le \epsilon_D, \qquad \mathrm{KL}(\hat{p}_w(r) \,\|\, \pi_R) \le \epsilon_R,$$

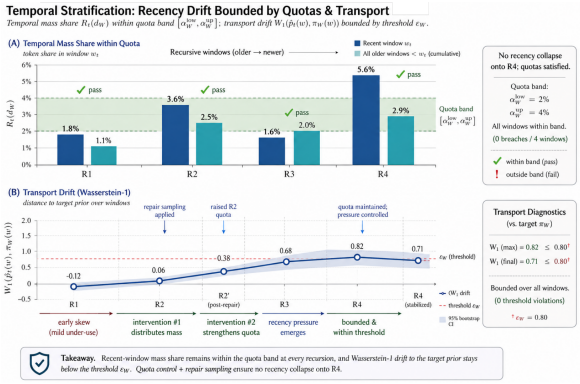


Figure 6: ***Temporal Stratification and Recency Drift (T3).*** ACP prevents recursive data mass from collapsing onto the newest synthesis window. ***Top:*** temporal quotas keep window-level token shares within the allowed band, avoiding over-concentration on R4. ***Bottom:*** Wasserstein drift $W_1(\hat{p}(w), \pi_W)$ remains below the pre-registered threshold ϵ_W, while repair sampling and quota updates bound recency drift across windows.

$$\mathrm{TV}(\hat{p}_w(d, r), \pi_D(d)\pi_R(r)) \le \epsilon_{DR}, \qquad W_1(\hat{p}_w(r), \pi_R) \le \epsilon_W.$$

Equivalently, T3 is the gate

$$g_{\mathrm{iso}}(\mathcal{A}_w) = \min\Big\{\epsilon_D - \mathrm{KL}_D,\ \epsilon_R - \mathrm{KL}_R,\ \epsilon_{DR} - \mathrm{TV}_{DR},\ \epsilon_W - W_1\Big\} \ge 0.$$

The joint term is important: marginal domain and time distributions may look balanced while a few domain–window cells dominate. The TV_{DR} constraint prevents such hidden bubbles, while W_1 penalizes recency drift by transport distance rather than only by bin counts. Thus, T3 prevents recursive gradients from over-concentrating on dominant styles or recent generations, reducing dataset-shift and concept-drift amplification [Quinonero-Candela et al., 2008; Gama et al., 2014]. Fig. 5 visualizes the domain–window coverage ratios controlled by T3, while Fig. 6 shows corresponding temporal quota and W_1-drift diagnostics.

2.4 T4: Red-Queen Critics

Motivation. Recursive collapse affects not only outputs, but also ***mechanisms***. Models can remain semantically correct while converging toward a single rhetorical scaffold, reasoning template, or teacher style. T4 therefore preserves ***plural reasoning routes*** through a ***mechanism-diverse teacher lattice*** and rotating critics, countering the broader degeneration and diversity-loss effects observed in recursive or repetitive generation regimes [Shumailov et al., 2024a; Deng et al., 2024b; Holtzman et al., 2020; Vijayakumar et al., 2018].

For teachers T_k, T_j, define symmetric route divergence over the accepted pool:

$$D^{k,j}_{\text{route}}(\mathcal{A}_w) = \mathbb{E}_{x \in \mathcal{A}_w}\Big[\mathrm{KL}(P_{T_k}(\cdot|x) \| P_{T_j}(\cdot|x))_{\text{sym}}\Big].$$

ACP aggregates route diversity across the teacher lattice:

$$D_{\text{route}}(\mathcal{A}_w) = \frac{2}{K(K-1)} \sum_{k<j} D^{k,j}_{\text{route}}(\mathcal{A}_w).$$

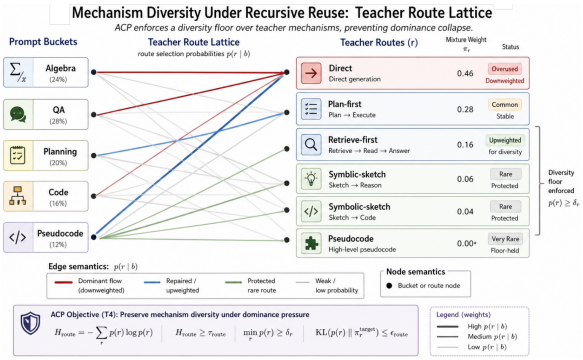


Figure 7: ***Red-Queen Critics (T4). ACP preserves plural reasoning mechanisms*** through a rotating teacher–critic lattice, preventing recursive training from collapsing onto a single dominant template. Routes include direct reasoning, chain-of-thought, retrieval-first decomposition, symbolic execution, and plan-first reasoning. Critics penalize ***route monoculture*** and ***stylistic sameness***, keeping multiple mechanisms available for the same semantic target.

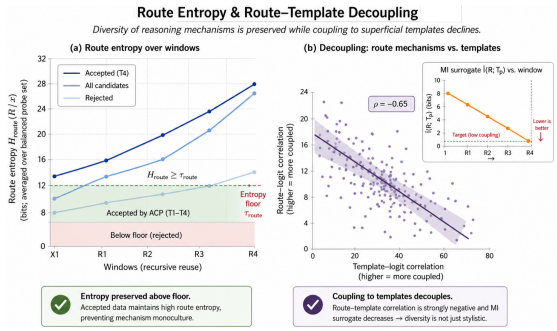


Figure 8: ***Route Entropy & Route–Template Decoupling (T4). Left:*** accepted samples maintain route entropy $H_{\text{route}}(R \mid x)$ above the floor τ_{route}, preventing mechanism monoculture across recursive windows. ***Right:*** route–template coupling declines: template-logit correlation is negatively associated with route-logit correlation, and the mutual-information surrogate $\widehat{I}(R; T_p)$ trends downward. Together, T4 preserves ***mechanism diversity*** rather than merely increasing superficial stylistic variation.

In parallel, let $p_w(r \mid x)$ denote the induced distribution over route families r. ACP monitors route entropy:

$$H_{\text{route}}(\mathcal{A}_w) = \mathbb{E}_{x \in \mathcal{A}_w}\left[-\sum_r p_w(r \mid x) \log p_w(r \mid x)\right].$$

T4 is then the gate

$$g_{\text{route}}(\mathcal{A}_w) = \min\Big\{D_{\text{route}}(\mathcal{A}_w) - \rho_{\min}, \\ H_{\text{route}}(\mathcal{A}_w) - \tau_{\text{route}}\Big\} \geq 0.$$

T4 preserves ***multiple valid solution mechanisms***: direct reasoning, chain-of-thought, retrieval-first decomposition, symbolic/pseudocode execution, plan-first reasoning, terse explanations, and discursive explanations. Rotating critics penalize ***template fixation***, ***stylistic sameness***, and ***route collapse***, sustaining mechanistic diversity under recursive reuse. Fig. 7 illustrates the ***teacher-route lattice*** used to preserve mechanism diversity, while Fig. 8 shows that ***route entropy remains above its floor*** and ***route–template coupling declines*** across recursive windows.

2.5 Tripwires, Adaptation, and Gate Interlock

ACP monitors a compact vector of collapse statistics:

$$\mathbf{m}(w) = [m_{\text{ent}}, m_{\text{tail}}, m_{\text{nov}}, m_{\text{iso}}, m_{\text{cal}}, m_{\text{route}}]_w,$$

covering entropy health, tail support, novelty growth, domain/time divergence, calibration, and route diversity. Let $\boldsymbol{\tau}$ denote the corresponding pre-registered thresholds. ACP triggers intervention only under ***persistent breach***:

$$\exists j: \quad m_j(w) > \tau_j \ \wedge \ m_j(w-1) > \tau_j.$$

This ***two-window rule*** prevents overreaction to stochastic fluctuations while still intervening before recursive drift compounds. When triggered, ACP adapts by tightening entropy bands, raising tail/novelty floors, rebalancing domains, reducing duplicate tolerance, or rotating critics. Calibration and hard-slice risk are tracked through ECE/Brier-style diagnostics and CVaR-NLL, respectively [Guo et al., 2017; Rockafellar and Uryasev, 2000].

The four gates are intentionally interdependent:

T1: learnability → T2: support → T3: coverage → T4: mechanisms.

T1 ensures that admitted data remain learnable; T2 preserves rare support and novelty; T3 distributes that support across domains and time; and T4 preserves plural reasoning mechanisms. Together, ACP converts recursive synthetic fine-tuning from uncontrolled self-feeding into ***auditable closed-loop control***.

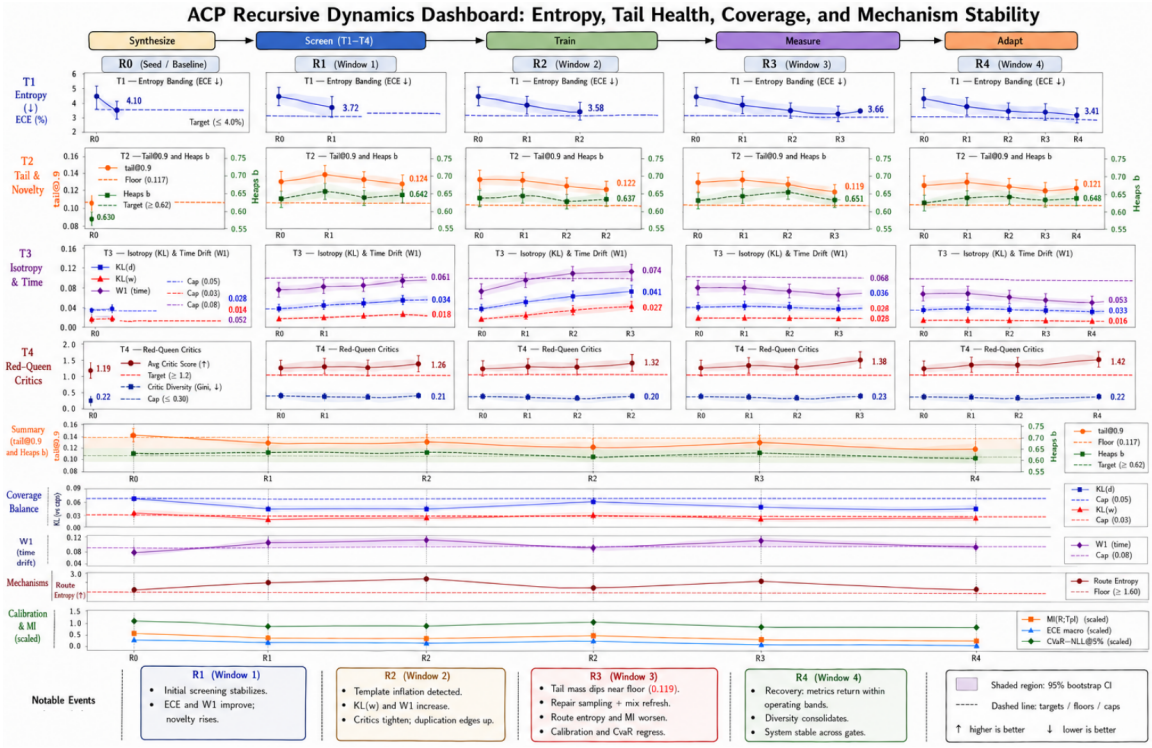


Figure 9: ***ACP prevents recursive synthetic collapse across four recursive windows.*** *Top band:* the closed-loop ACP pipeline per window—***Synthesize*** $\rightarrow$ ***Screen (T1–T4)*** $\rightarrow$ ***Train*** $\rightarrow$ ***Measure*** $\rightarrow$ ***Adapt***. *Middle grid:* columns correspond to ***R1–R4*** and rows correspond to ***T1–T4***, each rendered with unified row-wise scales and tripwire status indicators. ***T1*** tracks calibration stability via ***ECE*** and entropy band-pass rate; ***T2*** tracks ***tail@0.9***, ***Heaps'*** b, novelty growth, and duplicate-control caps; ***T3*** tracks ***KL*** imbalance and W_1 recency drift against isotropy caps; ***T4*** tracks ***route entropy*** and ***route–template coupling*** as mechanism-diversity diagnostics. *Bottom strips:* aggregate telemetry across $R1 \rightarrow R4$ shows ***stable tail support***, ***bounded coverage drift***, ***preserved reasoning plurality***, and ***controlled calibration / CVaR-NLL risk*** despite recursive reuse pressure.

3 Experiments and Results

We test ACP under ***recursive synthetic reuse*** across $R1 \rightarrow R4$. The goal is simple: ***does controlled admission prevent collapse before headline accuracy fails?*** We compare ACP against unconstrained Recursive-Self training and track the failure axes ACP is designed to protect: ***learnability***, ***tail support***, ***coverage balance***, ***calibration***, and ***mechanism diversity***, with pre-registered tripwires showing when drift becomes persistent rather than stochastic noise under sustained recursive reuse pressure.

Experimental Setup

Student. We train a ***Llama-3-8B*** student using ***LoRA*** and ***QLoRA***. ***Teacher lattice.*** Synthetic supervision is generated from ***Llama-3-70B***, ***Mixtral-8×22B***, and ***Qwen2-72B***. ***Recursive windows.*** We perform four recursive synthetic fine-tuning windows $R1 \rightarrow R4$ under matched accepted-token budgets (~ 50M/window). ***Routes.*** Generation routes include ***direct***, ***chain-of-thought***, ***plan-first***, ***retrieve-first***, and ***symbolic/pseudocode*** reasoning. ***Diagnostics.*** ACP monitors ***T1 learnability***, ***T2 tail support***, ***T3 domain/time drift***, and ***T4 mechanism diversity*** through entropy, calibration, Zipf-tail mass, Heaps-law growth, Wasserstein drift, route entropy, and route–template coupling.

3.1 Main Results

Fig. 9 summarizes the central empirical result: ***recursive synthetic fine-tuning does not inevitably collapse if data admission is regulated as a closed-loop control problem.*** Across recursive windows, ACP remains within pre-registered stability bounds for ***learnability***, ***support geometry***, ***coverage isotropy***, and ***mechanism diversity***. The result is best read in three parts.

(i) Collapse appears before headline accuracy fails. Under unconstrained Recursive-Self training, we observe the classical collapse trajectory predicted by prior work [Shumailov et al., 2024a; Deng et al., 2024b]: ***entropy compression***, ***duplicate amplification***, ***tail shrinkage***, ***style monoculture***, and ***calibration drift***. These failures emerge early (typically by $R2$ or $R3$), even when aggregate benchmark performance initially appears stable. Thus, collapse is first visible in the telemetry: tail mass weakens, duplicate pressure rises, coverage drifts, and calibration degrades before downstream averages fully expose the damage.

(ii) ACP controls the whole failure cascade, not one metric. Entropy band-pass rates remain above target thresholds, ECE stays bounded, Zipf-tail mass remains within tolerance bands, Heaps-law novelty growth increases across windows, and Wasserstein drift remains below intervention caps. Across $R1 \rightarrow R4$, ACP preserves tail@0.9 within the certified support band, maintains W_1 below the drift threshold, reduces ECE, and keeps route entropy above the diversity floor, whereas Recursive-Self breaches duplicate and calibration tripwires by $R2 \rightarrow R3$. This joint stability is important: ACP does not merely improve one axis while hiding degradation elsewhere; it keeps support, coverage, calibration, and mechanism metrics simultaneously inside the admissible region.

(iii) Stability does not come by collapsing diversity. Route entropy increases while route–template coupling falls, showing that recursive supervision preserves ***multiple reasoning mechanisms, not merely surface-level stylistic variation***. This matters because a model can remain semantically correct while silently converging to one dominant explanatory scaffold. ACP prevents this hidden mechanism collapse: the accepted data preserve plural routes while suppressing template fixation. At the same time, CVaR-NLL remains stable or improves across domains, and GLUE macro performance remains non-degrading throughout recursion. Thus, ACP does not buy stability by sacrificing utility; it converts recursive synthetic training into an auditable regime where ***tail health***, ***coverage balance***, ***calibration***, and ***reasoning plurality*** remain jointly measurable and controlled.

Collapse Certification

ACP certifies collapse only under persistent tripwire breaches. ***T1 (learnability).*** Entropy-band failure or calibration drift. ***T2 (support).*** Tail shrinkage, Heaps degradation, novelty loss, or duplicate inflation. ***T3 (coverage).*** Domain/time KL, Wasserstein drift, style monoculture, or recency lock-in. ***T4 (mechanisms).*** Route-entropy collapse or rising route–template coupling. Recursive-Self trips multiple bounds by $R2 \rightarrow R3$; ACP stays certified across $R1 \rightarrow R4$.

3.2 Ablations

Removing individual ACP gates reproduces their targeted failure modes and shows that no single constraint is sufficient to prevent recursive collapse. Eliminating ***T1*** removes the learnability filter: low-entropy rote samples and high-entropy noisy samples both enter training, causing unstable entropy growth, noisier gradients, and worsening calibration. Removing ***T2*** weakens support preservation; rare-token mass shrinks, near-duplicate templates expand, and Heaps-law growth flattens, indicating that apparent corpus size no longer corresponds to genuine support expansion.

Without ***T3***, the accepted pool remains locally plausible but becomes globally skewed. Recursive training progressively over-concentrates on dominant domains, recent synthesis windows, and stylistic formats, producing recency monoculture and elevated Wasserstein drift. Finally, removing ***T4*** preserves surface correctness for some tasks but collapses the route distribution: route entropy falls, route–template coupling rises, and generations converge toward a single rhetorical scaffold rather than maintaining multiple valid mechanisms.

These failures are complementary rather than redundant. ***T1*** controls whether examples are learnable, ***T2*** controls whether rare support survives, ***T3*** controls whether support remains balanced across domains and time, and ***T4*** controls whether solutions remain mechanistically plural. Together, the ablations support the central ACP hypothesis: ***recursive collapse is not a single phenomenon, but a coupled instability involving learnability, support geometry, coverage drift, and mechanistic diversity simultaneously.***

3.3 GLUE Stability Under Recursive Reuse

Using a Llama-3-8B student with LoRA/QLoRA adaptation and synthetic supervision from Llama-3-70B, Mixtral-8×22B, and Qwen2-72B teachers, ACP remains non-degrading on GLUE macro performance ($+0.2$ macro points across $R0 \rightarrow R4$) while preserving recursive stability.

GLUE Stability vs. Collapse Telemetry

GLUE stability. ACP remains non-degrading on GLUE ($+0.3\%$ across $R1 \rightarrow R4$); task-wise results appear in Appendix B.8. ***Telemetry.*** Recursive-Self shows entropy compression, tail shrinkage, recency drift, and route collapse by $R2 \rightarrow R3$, while ACP keeps all telemetry channels within bounds.

4 Conclusion

Synthetic repetition is not the enemy; uncontrolled repetition is. We introduced ***ACP***, a provenance-agnostic protocol regulating internal geometry via entropy banding, tail/novelty floors, domain isotropy, and Red-Queen critics. ACP preserves ***information heterozygosity***—health, coverage, calibration, and mechanism plurality—while ablations recover failures. Collapse is not inevitable; it is an auditable control problem.

Limitations

ACP should be read as a ***controlled anti-collapse protocol*** for recursive synthetic fine-tuning, not as a universal proof that synthetic recursion is safe. This distinction matters because recent work gives a nuanced picture: recursive replacement can induce collapse and tail loss [Shumailov et al., 2024b; Alemohammad et al., 2024; Dohmatob et al., 2024, 2025], while accumulation, verification, correction, and mixture control can mitigate collapse under specific assumptions [Gerstgrasser et al., 2025; Gillman et al., 2024; Feng et al., 2025]. ACP is positioned within this second line: it does not deny collapse risk; it makes the risk ***auditable, gated, and intervention-triggered***.

i) Controlled recursion, not frontier-scale pretraining. Our experiments study repeated fine-tuning with a Llama-3-8B student, LoRA/QLoRA adaptation, and a fixed recursive protocol. We do not claim that ACP fully solves trillion-token frontier pretraining, continuously refreshed web crawls, or multimodal self-consuming ecosystems. Instead, ACP tests a narrower claim: when recursion is measurable, collapse can be treated as a ***closed-loop distribution-control problem*** rather than as an unavoidable consequence of synthetic provenance.

ii) Finite recursive horizon. We evaluate four recursive windows, $R1 \to R4$, with approximately 50M accepted tokens per window. This is sufficient to reveal early collapse in Recursive-Self, but not sufficient to rule out slower failure modes. Longer chains may expose delayed tail erosion, creeping calibration drift, or gradual movement in $W_1(\hat{p}(w), \pi_W)$ beyond our observed range. Future work should stress-test ACP under $R8/R10$ recursion and under increasing synthetic-to-human ratios.

iii) Diagnostics are partly protocol-internal. ACP explicitly regulates entropy band-pass, Zipf-tail mass, novelty, duplicate rate, domain/time drift, and route entropy. Therefore, preserving H_{seq}, TailMass_q, Novelty_n, W_1, and H_{route} is not, by itself, proof of universal collapse prevention. Our claim is more specific: ACP prevents the ***measured collapse cascade*** predicted by recursive-collapse work—tail loss, duplication, calibration drift, support contraction, and mechanism monoculture. External tests such as rare-entity recall, multilingual tail preservation, long-tail factual QA, and hard-slice calibration remain necessary.

iv) Threshold sensitivity and control-policy design. ACP depends on explicit thresholds: entropy bands $[h_w^-, h_w^+]$, tail floors η_q, novelty floors γ_n, duplicate caps κ, drift bounds $\epsilon_D, \epsilon_R, \epsilon_{DR}, \epsilon_W$, and route-entropy floors τ_{route}. Although these thresholds are pre-registered, they may interact with domain skew, teacher quality, token budget, and recursive depth. ACP currently uses conservative hand-specified control rules; learning adaptive thresholds or deriving stability conditions for $g(A) \succeq 0$ remains open.

v) Teacher-lattice dependence. ACP uses a diverse teacher lattice: Llama-3-70B, Mixtral-8×22B, and Qwen2-72B. Some gains may therefore arise from teacher diversity rather than the admission protocol alone. This is not a flaw in ACP—teacher diversity is part of the design—but it limits causal attribution. Stronger separation requires single-teacher ACP, multi-teacher without gates, teacher-lattice-only, and gate-only baselines.

vi) Relation to accumulation, verification, and correction methods. Recent work suggests that collapse may be mitigated by accumulating real and synthetic data, applying verification, or correcting synthetic samples toward the true distribution. ACP is complementary: it does not assume access to an oracle corrector, a persistent clean human pool, or a binary synthetic detector. Instead, it controls admitted information geometry through learnability, tail support, coverage, and route plurality. However, combining ACP with accumulation, verifier-based selection, or grounding-aware correction may yield stronger guarantees than ACP alone.

vii) Mechanism diversity is approximate. T4 uses route families—direct, chain-of-thought, plan-first, retrieve-first, and symbolic/pseudocode—as operational proxies for mechanism plurality. These are not ground-truth decompositions of internal computation. Two surface-distinct routes may share latent mechanisms, while similar explanations may arise from different computations. Thus, H_{route} and route–template decoupling should be interpreted as ***auditable proxies***, not definitive evidence of internal reasoning diversity.

viii) Novelty is not automatically useful. T2 preserves tail mass and unseen n-grams, but novelty may be constructive or harmful. Rare generations can expand useful support, but they may also in-

troduce hallucinated facts, weak grounding, or unstable supervision. T1 reduces this risk by rejecting low-entropy rote text and high-entropy noise, but ACP does not fully solve the distinction between ***constructive novelty*** and ***destabilizing novelty***. Grounded novelty estimation remains a key extension.

ix) Collapse scope is deliberately bounded. ACP targets ***distributional collapse***: entropy compression, support shrinkage, duplication, coverage drift, calibration degradation, and route monoculture. It does not comprehensively address factual contamination, privacy leakage, harmful bias amplification, adversarial synthetic poisoning, reward hacking, or strategic manipulation. These risks may interact with collapse, but they require additional safeguards beyond the present protocol.

x) No claim of indefinite synthetic sustainability. ACP does not prove that synthetic recursion can continue indefinitely. It shows that, in a controlled recursive fine-tuning regime, collapse behaves less like a binary provenance problem and more like an ***auditable information-geometry instability***. The broader question—whether such stabilization survives frontier scale, uncontrolled open-web recursion, and multimodal synthetic ecosystems—remains open.

References

Epoch AI. 2023. When will we run out of data? forecasting the limits of public human text for training large language models. Technical report, Epoch.

Sina Alemohammad, Josue Casco-Rodriguez, Lorenzo Luzi, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkoohi, and Richard G. Baraniuk. 2024. Self-consuming generative models go mad. In *International Conference on Learning Representations*.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *FAccT*.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *ICML*.

Chunyuan Deng, Yilun Zhao, Yuzhao Heng, Yitong Li, Jiannan Cao, Xiangru Tang, and Arman Cohan. 2024a. Unveiling the spectrum of data contamination in language models: A survey from detection to remediation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 16030–16057. Association for Computational Linguistics.

Chunyuan Deng, Yilun Zhao, Yuzhao Heng, Yitong Li, Jiannan Cao, Xiangru Tang, and Arman Cohan. 2024b. Unveiling the spectrum of data contamination in language models: A survey from detection to remediation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 16030–16057. Association for Computational Linguistics.

Thomas G. Dietterich. 2000. Ensemble methods in machine learning. In *Multiple Classifier Systems (MCS 2000)*, volume 1857 of *Lecture Notes in Computer Science*, pages 1–15. Springer.

Elvis Dohmatob, Yunzhen Feng, Arjun Subramonian, and Julia Kempe. 2025. Strong model collapse. In *International Conference on Learning Representations*.

Elvis Dohmatob, Yunzhen Feng, Pu Yang, Francois Charton, and Julia Kempe. 2024. A tale of tails: Model collapse as a change of scaling laws. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*.

Yunzhen Feng, Elvis Dohmatob, Pu Yang, Francois Charton, and Julia Kempe. 2025. Beyond model collapse: Scaling up with synthesized data requires verification. In *International Conference on Learning Representations*.

João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. 2014. A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4):44:1–44:37.

Leo Gao, Stella Biderman, Sid Black, and et al. 2021. The pile: An 800gb dataset of diverse text for language modeling. In *NeurIPS Datasets and Benchmarks*.

Pascal Germain, Alexandre Lacasse, François Laviolette, Mario Marchand, and Jean-François Roy. 2015. Risk bounds for the majority vote: From a pac-bayesian analysis to a learning algorithm. *arXiv preprint arXiv:1503.06210*.

Matthias Gerstgrasser, Rylan Schaeffer, Apratim Dey, Rafael Rafailov, Henry Sleight, John Hughes, Tomasz Korbak, Rajashree Agrawal, Dhruv Pai, Andrey Gromov, Daniel A. Roberts, Diyi Yang, David L. Donoho, and Sanmi Koyejo. 2025. Is model collapse inevitable? breaking the curse of recursion by accumulating real and synthetic data. In *International Conference on Learning Representations*.

Nate Gillman, Michael Freeman, Daksh Aggarwal, Chia-Hong Hsu, Calvin Luo, Yonglong Tian, and Chen Sun. 2024. Self-correcting self-consuming loops for generative model training. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 15646–15677.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 27.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. 2017. On calibration of modern neural networks. In *ICML*.

Daniel L. Hartl and Andrew G. Clark. 1997. *Principles of Population Genetics*, 3 edition. Sinauer Associates, Sunderland, Massachusetts.

Harold S. Heaps. 1978. Information retrieval, computational and theoretical aspects. *Academic Press*.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. In *NeurIPS Deep Learning Workshop*. ArXiv:1503.02531.

Jordan Hoffmann, Sebastian Borgeaud, and et al. 2022. Training compute-optimal large language models. In *arXiv:2203.15556*.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2020. The curious case of neural text degeneration. In *ICLR*.

Nikhil Kandpal, Eric Wallace, and Colin Raffel. 2022. Deduplicating training data mitigates privacy risks in language models. In *NeurIPS*.

Jared Kaplan, Sam McCandlish, and et al. 2020. Scaling laws for neural language models. In *arXiv:2001.08361*.

Divyansh Kaushik, Eduard Hovy, and Zachary C. Lipton. 2020. Learning the difference that makes a difference with counterfactually-augmented data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 4078–4098.

M. Pawan Kumar, Benjamin Packer, and Daphne Koller. 2010. Self-paced learning for latent variable models. In *NeurIPS*.

Ludmila I. Kuncheva and Christopher J. Whitaker. 2003. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 51(2):181–207.

Clara Meister, Tiago Pimentel, Gian Wiher, and Ryan Cotterell. 2023. Locally typical sampling. *Transactions of the Association for Computational Linguistics*, 11:102–121.

Masatoshi Nei. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70(12):3321–3323.

Yixin Nie, Adina Williams, Emily Dinan, Mohit Bansal, Jason Weston, and Douwe Kiela. 2020. Adversarial NLI: A new benchmark for natural language understanding. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4885–4901. Association for Computational Linguistics.

Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. 2016. f-gan: Training generative neural samplers using variational divergence minimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 29.

Guilherme Penedo, Quentin Malartic, Daniel Hesslow, Ruxandra Cojocaru, Alessandro Cappelli, Hamza Alobeidli, Baptiste Pannier, Ebtesam Almazrouei, and Julien Launay. 2023. The refinedweb dataset for falcon llm: Outperforming curated corpora with web data, and web data only. *arXiv preprint arXiv:2306.01116*.

Joaquin Quinonero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D. Lawrence. 2008. Dataset shift in machine learning. *MIT Press*.

Colin Raffel, Noam Shazeer, Adam Roberts, and et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. In *JMLR*.

Jorma Rissanen. 1978. Modeling by shortest data description. *Automatica*, 14(5):465–471.

Jorma Rissanen. 1989. *Stochastic Complexity in Statistical Inquiry*. World Scientific.

R. Tyrrell Rockafellar and Stanislav Uryasev. 2000. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41.

Burr Settles. 2009. *Active Learning*. Synthesis Lectures on AI and ML.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. 2024a. Ai models collapse when trained on recursively generated data. *Nature*, 631:755–759.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. 2024b. Ai models collapse when trained on recursively generated data. *Nature*, 631:755–759.

Luca Soldaini, Rodney Kinney, Akshita Bhagia, Dustin Schwenk, David Atkinson, Russell Authur, Ben Bogin, Khyathi Chandu, Jennifer Dumas, Yanai Elazar, Valentin Hofmann, Ananya Harsh Jha, Sachin Kumar, Li Lucy, Xinxi Lyu, Nathan Lambert, Ian Magnusson, Jacob Morrison, Niklas Muennighoff, Aakanksha Naik, Crystal Nam, Matthew E. Peters, Abhilasha Ravichander, Kyle Richardson, Zejiang Shen, Emma Strubell, Nishant Subramani, Oyvind Tafjord, Pete Walsh, Luke Zettlemoyer, Noah A. Smith, Hannaneh Hajishirzi, Iz Beltagy, Dirk Groeneveld, Jesse Dodge, and Kyle Lo. 2024. Dolma: an open corpus of three trillion tokens for language model pretraining research.

Yixuan Su, Nigel Collier, and et al. 2022. A contrastive framework for neural text generation. In *NeurIPS*.

Leigh Van Valen. 1973. A new evolutionary law. *Evolutionary Theory*, 1:1–30.

Ashwin Vijayakumar, Michael Cogswell, and et al. 2018. Diverse beam search for improved description of complex scenes. In *AAAI*.

Pablo Villalobos, Anson Ho, Jaime Sevilla, Tamay Besiroglu, Lennart Heim, and Marius Hobbhahn. 2024. Will we run out of data? limits of llm scaling based on human-generated data.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2022. Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A. Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2023. Self-instruct: Aligning language models with self-generated instructions. *arXiv preprint arXiv:2212.10560*.

Maurice Weber, Daniel Y Fu, Quentin Gregory Anthony, Yonatan Oren, Shane Adams, Anton Alexandrov, Xiaozhong Lyu, Huu Nguyen, Xiaozhe Yao, Virginia Adams, Ben Athiwaratkun, Rahul Chalamala, Kezhen Chen, Max Ryabinin, Tri Dao, Percy Liang, Christopher Re, Irina Rish, and Ce Zhang. 2024. Redpajama: an open dataset for training large language models. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Sean Welleck, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. 2020. Neural text generation with unlikelihood training. In *International Conference on Learning Representations*.

Can Xu, Qingfeng Sun, Kai Zheng, Xiubo Geng, Pu Zhao, Jiazhan Feng, Chongyang Tao, Qingwei Lin, and Daxin Jiang. 2023. Wizardlm: Empowering large pre-trained language models to follow complex instructions. *arXiv preprint arXiv:2304.12244*.

George Kingsley Zipf. 1949. *Human Behavior and the Principle of Least Effort*. Addison-Wesley.

Frequently Asked Questions (FAQs)

✻ What is the strongest bounded claim of ACP?

➠ ACP does ***not*** claim that synthetic recursion is universally safe, nor that model collapse is solved at frontier scale. The central claim is narrower and more auditable:

> ***Recursive synthetic fine-tuning need not collapse immediately when admitted data are controlled along learnability, tail support, novelty, coverage, calibration, and mechanism-diversity axes. ACP operationalizes this as a closed-loop protocol, and in our controlled setting prevents the measured collapse cascade that appears under unconstrained Recursive-Self.***

This positioning is deliberately weaker than "synthetic data are safe," but stronger than a diagnostic-only study: ACP provides an ***intervention protocol*** with gates, tripwires, and adaptation.

✻ Does ACP merely preserve the diagnostics it was designed to preserve?

➠ No. We agree that preserving protocol-internal quantities alone would be insufficient. ACP is evaluated as a ***collapse-cascade intervention***, not as a single-score optimizer. The evidence has three parts:

- ***Pre-registered stability:*** ACP maintains collapse indicators across entropy, tail support, novelty, coverage, calibration, and route diversity.
- ***Gate-level causal attribution:*** removing T1/T2/T3/T4 recovers the corresponding predicted failures: degenerate entropy, tail shrinkage, style/recency bubbles, and route monoculture.
- ***External non-degradation:*** GLUE macro remains non-degrading while Recursive-Self breaches duplicate and calibration tripwires.

Thus, the claim is not that ACP optimizes its own metrics, but that it interrupts the ***coupled failure pattern*** expected under recursive self-feeding.

✻ Why is ACP not just another data filter?

➠ ACP is not a static filter; it is a ***recursive closed-loop control protocol***. A static filter accepts or rejects samples once. ACP repeatedly synthesizes, screens, trains, measures, and adapts:

$$\textsc{Synthesize} \rightarrow \textsc{Screen} \rightarrow \textsc{Train} \rightarrow \textsc{Measure} \rightarrow \textsc{Adapt}.$$

This distinction matters because collapse is temporally compounded. The question is not only which samples enter training at window w, but how measurements after training S_w alter the next accepted pool A_{w+1}. ACP therefore treats recursion as a ***controlled data-generation process***, rather than as one-shot quality filtering.

✻ How is ACP different from provenance filtering, AI-text detection, or human-only filtering?

➠ ACP rejects the premise that provenance is the right control variable. A synthetic paragraph can contain useful novelty; a human paragraph can be duplicated boilerplate. In mixed future crawls, "AI vs. human" labels are also noisy, gameable, and increasingly unavailable. ACP instead controls the ***geometry of admitted information***:

- **T1:** learnability via entropy banding;
- **T2:** support via tail mass, novelty, and duplicate caps;
- **T3:** coverage via domain/time isotropy;
- **T4:** mechanism plurality via Red-Queen critics.

The goal is not to detect origin; it is to preserve the statistical conditions under which recursive learning remains stable.

✻ Is four recursive windows enough to prove collapse prevention?

➠ No. Four windows are not enough to rule out all delayed collapse modes. We interpret $R1 \rightarrow R4$ as an ***early-collapse stress test***, not as proof of asymptotic stability. The key observation is that, under matched budgets,

Recursive-Self already shows collapse signatures by $R2$–$R3$, while ACP remains within pre-registered tripwires through $R4$.

Longer horizons may reveal slower failures: delayed tail erosion, creeping calibration drift, or gradual movement in $W_1(\hat{p}(w), \pi_W)$. A natural extension is $R8/R10$ recursion, possibly with smaller per-window budgets, to test whether ACP delays or prevents long-horizon collapse.

✻ Are the gains simply due to the diverse teacher lattice?

➠ Teacher diversity is part of ACP, but it does not fully explain the result. The key comparison is not *diverse teachers vs. weak teachers*; it is ***diverse generation without admission control*** vs. ***diverse generation with recursive gates and tripwires***. ACP uses the same teacher-lattice principle throughout, but gate removals still recover the expected collapse modes.

That said, teacher-lattice dependence is a real limitation. Stronger causal separation should include:

- **single-teacher ACP**;
- **multi-teacher without gates**;
- **teacher-lattice-only baseline**;
- **gate-only baseline with weaker generators**.

These would separate ***source diversity*** from ***admission-control diversity***.

✻ Are ACP thresholds hand-tuned?

➠ The current thresholds are pre-registered and conservative, but we do not claim they are theoretically optimal. ACP exposes these thresholds explicitly:

$$[h_w^-, h_w^+], \quad \eta_q, \quad \gamma_n, \quad \kappa, \quad \epsilon_D, \epsilon_R, \epsilon_{DR}, \epsilon_W, \quad \tau_{\text{route}}.$$

This makes stability ***auditable*** rather than hidden inside an opaque quality score. The limitation is that threshold selection may interact with domain skew, teacher quality, token budget, and recursive depth. A stronger future version would learn adaptive thresholds or derive stability conditions for

$$g(A) \succeq 0.$$

✻ How sensitive is ACP to threshold choices?

➠ ACP should be stress-tested through threshold sweeps. The most important sensitivity plots are:

- entropy-band width $[h_w^-, h_w^+]$ vs. ECE and band-pass rate;
- tail floor η_q vs. rare-token retention and downstream utility;
- novelty floor γ_n vs. hallucination/noise rate;
- duplicate cap κ vs. repetition and data efficiency;
- route floor τ_{route} vs. route diversity and answer quality.

The expected healthy pattern is a ***stable middle regime***: too-loose thresholds recover Recursive-Self-like drift, while too-tight thresholds reduce usable data or admit overly artificial diversity.

✻ Why not simply use deduplication, quality filtering, or mixture rebalancing?

➠ These methods are useful but incomplete. Deduplication controls clone inflation but not entropy quality, temporal drift, or route monoculture. Quality filtering removes low-quality samples but may over-select head-distribution templates. Mixture rebalancing improves coverage but does not guarantee tail novelty or mechanism diversity.

ACP combines these concerns into one interlocking protocol:

$$\text{T1: learnability} \rightarrow \text{T2: support} \rightarrow \text{T3: coverage} \rightarrow \text{T4: mechanisms.}$$

The point is not that existing hygiene methods are wrong, but that recursive collapse is ***multi-axis*** and therefore requires coordinated control.

✻ How does ACP relate to recent work claiming model collapse is inevitable or avoidable?

➠ ACP sits between the two views. Collapse studies show that recursive replacement can induce tail loss, distributional narrowing, and degeneration. Other work shows that accumulation, correction, verification, or controlled

mixtures can mitigate collapse under specific assumptions. ACP adopts the second interpretation: collapse is not denied, but treated as ***measurable and controllable***.

The distinction is:

- ***Inevitable-collapse framing:*** recursive self-feeding depletes tails and amplifies model errors.
- ***ACP framing:*** recursive training becomes dangerous when admission is uncontrolled; with explicit gates and tripwires, the failure cascade can be delayed or interrupted in controlled regimes.

Thus ACP does not refute collapse theory; it operationalizes a mitigation protocol for the collapse mechanisms that theory identifies.

✻ Does ACP prove that synthetic data can replace human data?

➠ No. ACP is not a replacement argument. It does not claim that synthetic data are equivalent to human data, nor that human data can be removed from future training. Instead, ACP addresses the more realistic setting where synthetic data are increasingly present in training mixtures, either intentionally or through contaminated web crawls.

The goal is to make recursive synthetic reuse ***safer, auditable, and less collapse-prone***, not to eliminate the need for human-authored, grounded, diverse, and domain-expert data.

✻ Does preserving novelty risk admitting hallucinations?

➠ Yes. Novelty is not automatically useful. Rare generations may expand support, but they may also introduce hallucinated facts, weak grounding, or unstable supervision. ACP mitigates this by coupling novelty with other gates:

- **T1** rejects extremely high-entropy noise;
- **T2** requires novelty without duplicate inflation;
- **T3** prevents novelty from concentrating in one domain/window;
- **T4** discourages single-route synthetic artifacts.

However, ACP does not fully solve factual grounding. Retrieval verification, verifier-based correction, and grounding-aware novelty estimation are natural extensions.

✻ Does route entropy really measure mechanism diversity?

➠ Route entropy is a proxy, not a ground-truth measurement of internal computation. T4 does not claim to recover latent reasoning mechanisms exactly. Its purpose is to prevent recursive supervision from collapsing onto a single prompt scaffold, explanation style, or teacher route.

This is why ACP also tracks route–template decoupling: if route diversity were merely superficial wording variation, route labels would remain tightly coupled to templates. The current evidence supports ***auditable mechanism plurality***, but stronger validation would require representation-level route probes, human audits, or causal interventions on hidden states.

✻ Can ACP handle adversarial or malicious synthetic data?

➠ Only partially. ACP can suppress some adversarial effects indirectly if they appear as entropy outliers, duplicate bursts, domain/time drift, calibration degradation, or route collapse. However, ACP is not a complete defense against targeted poisoning, hidden triggers, privacy leakage, or strategically crafted synthetic contamination.

A stronger adversarial version would combine ACP with:

- contamination and near-duplicate audits;
- retrieval-grounded factual verification;
- adversarial red-team generation;
- provenance and watermark signals where available;
- robust training or poisoning-detection methods.

ACP should therefore be viewed as a ***collapse-control layer***, not a full security layer.

✻ What external evaluations would most strengthen ACP?

➠ The most valuable additions are evaluations where success cannot be reduced to ACP's own gates:

- **Rare-entity factual recall:** tests whether tail preservation improves knowledge retention.
- **Multilingual and dialectal tail slices:** tests minority-language and low-resource support.
- **Long-tail QA / hard-slice calibration:** tests whether calibration gains hold under difficult slices.
- **Contaminated-crawl simulation:** mixes human, synthetic, duplicated, and degraded generations.
- **Longer recursion:** $R8/R10$ windows to test delayed collapse.

These are natural next-step stress tests rather than requirements for the narrower controlled-regime claim.

✻ What are the main computational costs of ACP?

➠ ACP adds cost at three points:

- **screening cost:** entropy scoring, novelty checks, duplicate detection, and domain/time accounting;
- **teacher-lattice cost:** generating and comparing multiple routes or teachers;
- **diagnostic cost:** calibration, tail statistics, W_1 drift, and route-template coupling.

However, ACP is designed for ***accepted-token efficiency***: it spends extra computation before training to avoid wasting gradient budget on collapse-inducing data. The relevant cost comparison is therefore not only wall-clock screening overhead, but whether ACP reduces downstream retraining, failed recursive runs, or late-stage data repair.

✻ Can ACP be combined with accumulation, verification, or correction methods?

➠ Yes. ACP is complementary to these methods. Accumulation methods preserve a clean or partially clean data pool; verification methods reject low-quality synthetic samples; correction methods move synthetic data closer to target distributions. ACP supplies the recursive control layer:

$$\text{accumulate / verify / correct} \quad + \quad \text{ACP gates and tripwires.}$$

In practice, the strongest system may combine all of them: preserve human data where available, verify factuality, correct synthetic errors, and use ACP to maintain entropy, tail support, coverage, and mechanism plurality across recursive windows.

✻ What would falsify ACP's central claim?

➠ ACP would be weakened if a matched Recursive-Self or standard hygiene baseline preserves the same tail, calibration, novelty, coverage, and route-diversity profile without ACP gates. It would also be weakened if ACP maintains internal diagnostics but fails external long-tail tasks, rare-entity recall, multilingual tail slices, or hard-slice calibration.

A clean falsification test is:

> Run Recursive-Self, hygiene-only, teacher-lattice-only, and ACP under identical token budgets across $R8/R10$ windows; evaluate both internal diagnostics and external tail-sensitive tasks. If ACP does not separate from these baselines, the protocol is not adding meaningful collapse control.

✻ Why is "heterozygosity" the right metaphor?

➠ We use ***heterozygosity*** as an analogy for preserving multiple viable variants rather than allowing clonal dominance. In ACP, the variants are not biological alleles but informational modes: rare tokens, unseen n-grams, domain/time coverage, calibration behavior, and reasoning routes.

The metaphor is useful because recursive collapse is not only accuracy loss. A model can preserve headline performance while losing tail support, stylistic diversity, or alternative mechanisms. ACP's goal is therefore to preserve ***information heterozygosity***: multiple styles, domains, tails, and reasoning paths remain alive under recursion.

✻ **How should readers interpret "collapse certification"?**

➠ "Collapse certification" does not mean a formal proof that collapse is impossible. It means that collapse is declared only under ***pre-registered persistent tripwire breaches***. ACP monitors

$$m(w) = [m_{\text{ent}}, m_{\text{tail}}, m_{\text{nov}}, m_{\text{iso}}, m_{\text{cal}}, m_{\text{route}}]_w,$$

and intervenes only if a metric breaches its threshold across two consecutive windows:

$$\exists j : \; m_j(w) > \tau_j \;\wedge\; m_j(w-1) > \tau_j.$$

This prevents overreaction to stochastic noise while still catching recursive drift before it compounds.

✻ **One-sentence summary.**

➠ ***ACP does not claim that synthetic recursion is harmless; it shows that, in a controlled fine-tuning regime, the collapse cascade can be made measurable, gated, and interruptible through explicit control of learnability, support, coverage, calibration, and mechanism diversity.***

Appendix Contents

A The Anti-Collapse Protocol (ACP) . 20

A.1 Provenance–Agnostic Control via Information Geometry 20

A.2 Technique 1: Entropy Banding . . . 21

A.3 Technique 2: Tail–Mass & Novelty Floors with Strict Near–Duplicate Caps 28

A.4 Technique 3: Domain Isotropy with Temporal Stratification 32

A.5 Technique 4: Red–Queen Critics over a Mechanism–Diverse Teacher Lattice 36

B Experiments & Results 40

B.1 Data: Fixed Human Seed Slice & Synthetic Windows 41

B.2 Global Screening Outcomes and Data Health 44

B.3 Main Model Outcomes 44

B.4 Ablations and Baselines 44

B.5 Four Key Visual Summaries 45

B.6 Result Highlights with Pre–Registered Tripwires 45

B.7 Reproducibility Checklist 45

B.8 GLUE Utility & Collapse Certification 45

B.9 Cross-Lingual and Cross-Backbone Reproducibility (Reduced ACP) 50

A The Anti-Collapse Protocol (ACP)

Section goal. We motivate a *provenance–agnostic, information–geometry*–driven protocol for preventing collapse during *synthetic, repeated fine-tuning*. Rather than policing authorship ("AI vs. human"), ACP regulates the *shape* of information that enters training. Concretely, it imposes **acceptance constraints** on entropy, tails, novelty, domain/time coverage, and mechanism diversity; couples these to **auditable diagnostics** with **pre-registered tripwires**; and runs as a closed loop: **Synthesize** → **Screen** (T1–T4) → **Train** → **Measure** → **Adapt** (Fig. F0). This stance aligns with evidence that degeneration is caused by *recursive exposure to impoverished signal*—not by provenance per se [Shumailov et al., 2024a; Deng et al., 2024b; Bender et al., 2021].

Inputs assumed. *Student* $\mathcal{S}$: Llama-3–8B (base) with LoRA/QLoRA; *teacher lattice* $\{T_k\}_{k=1}^{K}$: Llama-3–70B, Mixtral 8×22B, Qwen2–72B; *windows* $W{=}4$ synthetic increments (R1→R4); *seed human slice* fixed across windows for calibration auditing.

Common notation (mini callout). Windows $w \in \{1..R\}$; domains $d \in \{1..D\}$; batch B; (Shannon) entropy H; token surprisal $\mathrm{sp}(t){=}-\log p(t)$; Zipf rank $r(t)$; tail set $\mathcal{T}_q{=}\{t : r(t) > \lfloor(1{-}q)|\mathcal{V}|\rfloor\}$; novelty set $\mathcal{G}_n$ of previously *unseen* n-grams; style clusters s; rotating critics C_t; candidate pool $\mathcal{C}$; acceptance set $\mathcal{A} \subset \mathcal{C}$. Empirical histograms are denoted $\hat{p}$; priors by π.

Cross-figure. F0 (flow): Synthesize → Screen (T1–T4) → Train → Measure → Adapt. F0 annotates where diagnostics and tripwires reside (bandpass, tail/novelty, KL-to-priors, route divergence, CVaR-NLL, ECE/Brier).

A.1 Provenance–agnostic control via information geometry

Consider a generative model $\mathcal{M}$ with token distribution $p_{\mathcal{M}}$ over vocabulary $\mathcal{V}$. In repeated synthetic fine-tuning, failure manifests as four coupled pathologies: *(i) tail death*—probability mass drifts from rare to frequent types; *(ii) support narrowing*—unique-type growth stalls despite more tokens; *(iii) style ossification*—a small set of rhetorical templates dominates; and *(iv) miscalibration*—the model becomes confidently wrong on hard slices [Shumailov et al., 2024a; Zipf, 1949; Heaps, 1978; Guo et al., 2017]. Read through a common geometric lens, training pushes $p_{\mathcal{M}}$ along directions that locally *reduce uncertainty* without introducing commensurate *new support*, thereby contracting the effective *information volume* explored by the model. What collapses is not just accuracy in the tails but the *capacity of the data pipeline* to feed gradients that **add** support rather than **redistribute** it toward the head.

***A minimal mathematical framing* (unnumbered).** For a sequence $x{=}(t_1,\ldots,t_{|x|})$ with conditionals $p(t_i \mid x_{<i})$, define the per-token entropy and mean surprisal

$$H_{\text{seq}}(x) \;=\; -\frac{1}{|x|}\sum_{i=1}^{|x|}\log p(t_i \mid x_{<i}), \qquad \overline{\mathrm{sp}}(x) \;=\; \frac{1}{|x|}\sum_{i=1}^{|x|} -\log p(t_i \mid x_{<i}).$$

Let the Zipf tail $\mathcal{T}_q$ contain the least frequent q-fraction of types. For any x define

$$\mathrm{TailMass}_q(x) \;=\; \frac{1}{|x|}\sum_{t\in x}\mathbf{1}[t \in \mathcal{T}_q], \qquad \mathrm{Novelty}_n(x) \;=\; \frac{1}{N_n(x)}\sum_{\gamma\in\mathrm{ngrams}_n(x)}\mathbf{1}[\gamma \notin \mathcal{G}_n],$$

where $N_n(x)$ counts n-grams. In healthy corpora, vocabulary growth follows *Heaps' law* $V(N) \approx K\,N^{\,b}$ with $b \in (0,1)$; collapse corresponds to $b\downarrow$ even as $N\uparrow$ [Heaps, 1978]. On calibration, if $\hat{p}(y \mid x)$ is the model's predictive confidence, reliability curves flatten under collapse and *expected calibration error* (ECE) increases [Guo et al., 2017]. Together, these metrics quantify how much *fresh support* enters training and how much *useful uncertainty* remains.

Why authorship is the wrong control knob. In modern crawls, the distinction "AI vs. human" is *noisy*, adversarially gameable, and orthogonal to informativeness [Bender et al., 2021]. A synthetic paragraph can carry **high tail mass** and **genuine novelty**; a human paragraph can be templated boilerplate. Conversely, a large fraction of web text is already LLM-assisted, so provenance filtering is both brittle and incomplete. What matters is whether incoming batches (i) inhabit a *learnable mid-entropy band* (to avoid rote and noise), (ii) sustain *tail and novelty throughput* (to keep rare modes alive), (iii) maintain *domain/time isotropy* (to prevent style/recency bubbles), and (iv) preserve *mechanism diversity* (so there remain many correct routes). These are the four signals ACP regulates.

A closed-loop, geometry-aware policy. Let $\mathcal{A}_w \subset \mathcal{C}_w$ be accepted samples in window w.

ACP enforces mid-entropy learnability by admitting batches with

$$\Pr_{x\sim B}[H_{\text{seq}}(x) \in \mathcal{B}_w] \;\geq\; \theta_H,$$

for a band $\mathcal{B}_w$ built from probe quantiles (refreshable as the student improves). It guarantees genuine support expansion with window-level floors

$$\text{TailMass}_q(\mathcal{A}_w) \;\geq\; \eta, \qquad \text{Novelty}_n(\mathcal{A}_w) \;\geq\; \nu,$$

while capping lexical/semantic/template duplicates to stop paraphrastic inflation (lexical shingle Jaccard, embedding cosine, per-template quotas). It balances coverage against priors $\pi(d)$ and $\text{Unif}(w)$ via divergence caps

$$\text{KL}(\hat{p}_w(d) \,\|\, \pi(d)) \leq \varepsilon_d, \quad \text{KL}(\hat{p}_w(w') \,\|\, \text{Unif}(w')) \leq \varepsilon_w, \quad \text{KL}(\hat{p}_w(d,w') \,\|\, \pi(d)\text{Unif}(w')) \leq \varepsilon_{dw},$$

and monitors style–time dependence $I(s;w)$ to detect *recency bubbles*. Finally, it preserves *plural reasoning routes* via a teacher lattice with route divergence

$$D_{\text{route}}(k,j) \;=\; \mathbb{E}_x\Big[\text{KL}(P_{T_k}(\cdot \mid x) \,\|\, P_{T_j}(\cdot \mid x))_{\text{sym}}\Big],$$

and *rotating critics* C_t (style/dup/confidence/consistency) that penalize sameness. Training proceeds, then *measurement*: we track **Heaps exponent** b, **tail/novelty**, **repetition/burstiness**, **coverage** KLs, **CVaR–NLL** on hard slices, and **ECE** [Rockafellar and Uryasev, 2000; Guo et al., 2017]. If *two consecutive windows* breach pre-registered thresholds, ACP *adapts*: tighten floors/caps, rotate critics, refresh bands, or reweight domains—*before* any further scaling.

Interlocking safeguards (why the four pieces together). *Entropy banding* (T1) prevents the system from drifting into a head-only regime (left-tail) or accepting incoherent junk (right-tail), ensuring that gradients are spent on *useful novelty*. *Tail and novelty floors with near-duplicate caps* (T2) convert that novelty into *guaranteed tail throughput* per batch and per window, while blocking token inflation via paraphrastic clones; this stabilizes Heaps-b and curbs repetition/burstiness. *Domain isotropy with temporal stratification* (T3) spreads those gains across *where* and *when* data comes from, suppressing style/recency bubbles and stopping any single basin from hijacking the gradient field. *Mechanism diversity with Red–Queen critics* (T4) keeps *how* answers are produced plural—direct vs. CoT, retrieval-first vs. plan-first, terse vs. discursive—so improvements do not collapse onto a single rhetorical or algorithmic scaffold. Together these gates implement a **closed-loop control** that directly counters the failure modes predicted by recursion theory and observed empirically: tail death, support narrowing, stylistic monoculture, and miscalibration [Shumailov et al., 2024a; Deng et al., 2024b].

Why now (necessity and urgency). *Scaling theory* indicates that gains at fixed compute come from **informative tokens**, not raw volume [Kaplan et al., 2020; Hoffmann et al., 2022]. *Recursion studies* demonstrate that unconstrained generated-data training erodes exactly those properties, accelerating collapse in the tails and drifting confidence calibration [Shumailov et al., 2024a]. Meanwhile, the web is already a *mixed ecosystem*; provenance labels are fading in reliability [Bender et al., 2021]. A **provenance-agnostic**, **geometry-aware** control law is therefore *urgent*: it **preserves heterozygosity** (rare knowledge, plural styles, diverse mechanisms) while allowing synthetic pipelines to grow responsibly. ACP turns a looming liability of synthetic data into an asset by governing the *geometry* of what we admit, not the *origin* of who wrote it.

Figure F0 (referenced above) illustrates the ACP loop and annotates the placement of diagnostics and tripwires across **Synthesize**, **Screen** (T1–T4), **Train**, **Measure**, and **Adapt**.

A.2 Technique 1: Entropy Banding — Admit Only Mid–Difficulty, Learnable Batches

Intuition and goal. In recursive synthetic pipelines, the learner tends to waste gradient on two regimes that catalyze collapse: *(i) head rote*—spans that are trivially predictable given the student's current hypothesis; and *(ii) degenerate junk*—spans whose uncertainty reflects noise rather than structure. The former *sharpens* already frequent templates, starving the tail; the latter injects variance without adding support, degrading calibration and stability. **Entropy Banding** institutes a provenance–agnostic control law: *admit only mid–difficulty, learnable batches*. Concretely, we accept a batch if a large fraction of its sequences lie in a *useful middle band* of predictive uncertainty, optionally coupled with span–level checks that prune hidden rote or junk. This pol-

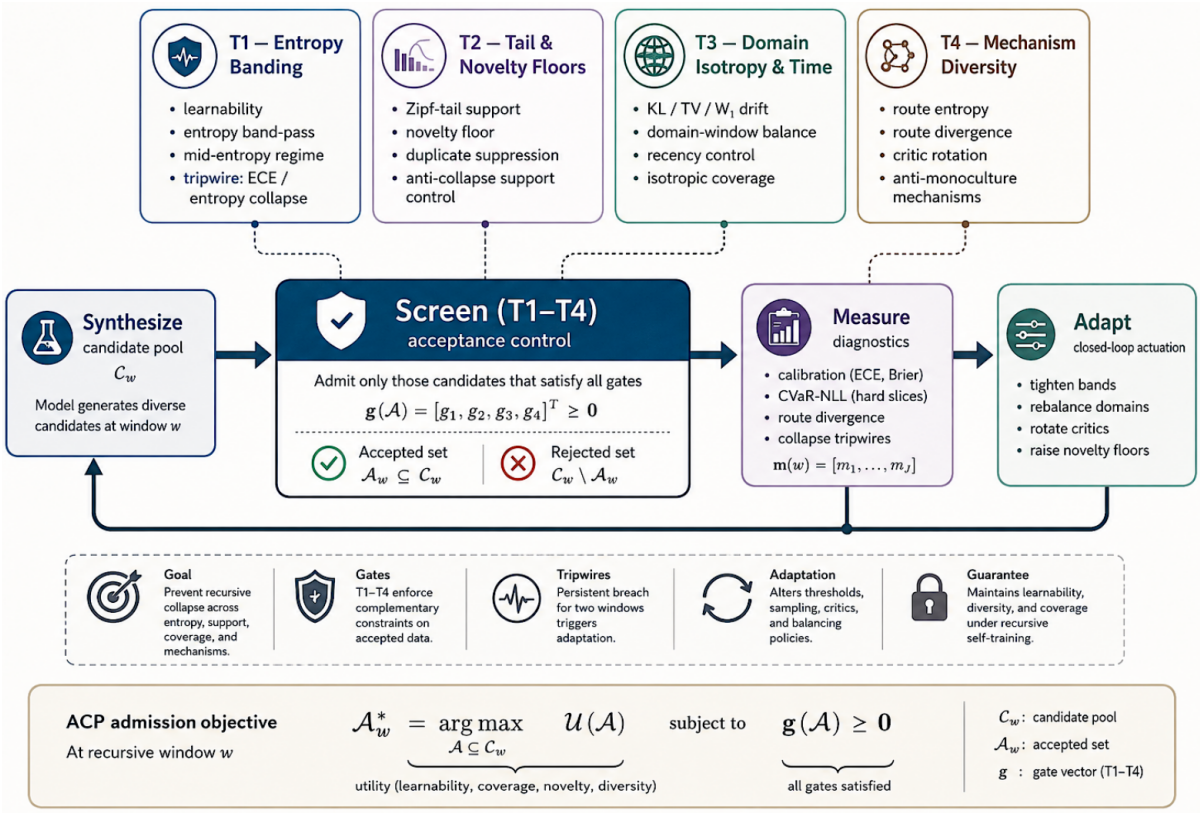


Figure 10: **ACP—Closed-Loop Overview (F0).** The figure summarizes the *Anti-Collapse Protocol* as a measurement-driven control loop: **Synthesize** → **Screen (T1–T4)** → **Train** → **Measure** → **Adapt**. Screening produces an *acceptance set* (**Accf–C**) used for training; measurement feeds back *tripwires* that trigger adaptive actions before the next window. **Left: Synthesize.** Candidate pool $\mathcal{C}$ is generated (teacher lattice, counterfactual recompositions, etc.). No provenance filters are applied; ACP is *origin-agnostic*. **Center: Screen (T1–T4).** Four interoperable gates regulate the *geometry of information: T1 Entropy Banding* keeps learning in the useful mid-entropy region (*band-pass %*, median H_{seq}; **Tripwire:** band-pass $< \theta_H$ or ECE ↑). *T2 Tail & Novelty Floors* guarantee rare-signal throughput (Zipf-tail mass, unique n-grams, Heaps-b; **Tripwire:** tail/novelty breach, $b \downarrow$ with repetition/burstiness ↑; strict near-duplicate caps). *T3 Domain Isotropy & Time* prevents style/recency lock-in (KL/TV/W divergences to $\pi(d)$ and Unif(w); style–time MI $I(s; w)$; **Tripwire:** KL caps breached or $I(s; w)$ spike). *T4 Mechanism Diversity* preserves plural reasoning routes (route divergence D_{route}, MI(teacher; style), critic pass-rate; **Tripwire:** divergence ↓, MI ↑, critic saturation). The accepted subset $\mathcal{A} \subset \mathcal{C}$ proceeds to training. **Right: Measure.** Global diagnostics include *Heaps-b*, tail/novelty, repetition/burstiness, domain/time KL, route divergence, *CVaR–NLL* on hard slices, and *ECE/Brier*. A *two-window* breach on any pre-registered metric escalates to Adapt. **Bottom: Adapt.** On tripwire, ACP *tightens floors/caps*, *rotates critics*, *refreshes entropy bands*, or *reweights domains/time*—closing the loop before the next synthesis window. *Abbreviations:* KL = Kullback–Leibler; TV = Total Variation; W = Wasserstein; ECE = Expected Calibration Error; CVaR = Conditional Value-at-Risk.

icy is the information–theoretic cousin of curriculum/self–paced learning [Bengio et al., 2009; Kumar et al., 2010], an MDL–aligned preference for *maximally compressible yet not already compressed* signal [Rissanen, 1978, 1989], and consistent with compute–optimal scaling where *informative tokens*, not raw volume, drive gains [Kaplan et al., 2020; Hoffmann et al., 2022].

Objects and probes. For a sequence $x = (t_1, \ldots, t_{|x|})$ with conditionals $p(t_i \mid x_{<i})$ under a probe model (student–on–policy or a stable reference), define per–token entropy and mean surprisal

$$H_{\text{seq}}(x) = -\frac{1}{|x|}\sum_{i=1}^{|x|} \log p(t_i \mid x_{<i}), \qquad \overline{\text{sp}}(x) = \frac{1}{|x|}\sum_{i=1}^{|x|} \left(-\log p(t_i \mid x_{<i})\right).$$

For *span–level* analysis, compute $H_{\text{span}}(x; \ell)$ over sliding windows of length ℓ. We estimate the distribution of H_{seq} (and optionally the joint $(H_{\text{seq}}, \overline{\text{sp}})$) on a refreshable *probe set* $\mathcal{P}$ held out from training.

The band and the acceptance rule. Let $\underline{h}$ and $\overline{h}$ be α and $(1-\alpha)$ quantiles of H_{seq} on $\mathcal{P}$; set the entropy band

$$\mathcal{B} = [\underline{h},\, \overline{h}] \quad (\textit{sequence–entropy band}).$$

Optionally, define a *joint* band $\mathcal{B}_{2\text{D}}$ as the highest–density contour of a KDE over $(H_{\text{seq}}, \overline{\text{sp}})$. For a batch B, the *band–pass* rate is

$$\text{BP}(B) = \Pr_{x \in B}[\, H_{\text{seq}}(x) \in \mathcal{B}\,].$$

We **accept** B iff $\text{BP}(B) \geq \theta_H$ (e.g., 90%). To block pockets of rote/junk embedded in otherwise mid–entropy sequences, we impose span–level guards:

$$\Pr_{s \subset x}[\, H_{\text{span}}(s;\ell) \leq \underline{h}_{\text{span}}\,] \leq \rho_{\min}, \qquad \Pr_{s \subset x}[\, H_{\text{span}}(s;\ell) \geq \overline{h}_{\text{span}}\,] \leq \rho_{\max},$$

with small tolerances $\rho_{\min}, \rho_{\max}$ (e.g., 5%). Bands are *frozen within a window* to avoid nonstationarity and *refreshed* at the next window with drift/width caps to preserve diversity.

Deeper intuition: what the band actually does to gradients. A useful way to visualize Entropy Banding is to imagine three "basins" of uncertainty. In the *left basin* (very low entropy), the student already assigns near–deterministic probabilities to upcoming tokens; gradients here are tiny and mostly *co-linear* with directions that amplify head templates (rote recollection). In the *right basin* (very high entropy), the student is effectively guessing; gradients are large but *misaligned* and dominated by noise or inconsistent contexts. The *middle basin* contains sequences where the model has a partial hypothesis—*uncertain, but for principled reasons*. Banding clips the left and right basins, reallocating gradient budget to the middle basin where updates (i) add *discriminative structure* and (ii) increase the model's *effective support* rather than merely sharpening head mass.

Formally, let $g(x) = \nabla \ell(\theta; x)$ under token–level NLL and denote the *true* risk gradient by $G = \nabla \mathcal{L}(\theta)$. Two statistics govern progress per step: *alignment* $\mathbb{E}[\langle G, g(x)\rangle]$ and *variance* $\mathbb{E}[\|g(x)\|^2] - \|\mathbb{E}[g(x)]\|^2$. The left basin hurts the first term (diminishing returns); the right basin explodes the second (instability). The middle band maximizes the alignment–variance tradeoff, a point long noted in *uncertainty sampling* and *expected model change* perspectives of active learning [Settles, 2009] and implicitly in self–paced curricula [Bengio et al., 2009; Kumar et al., 2010]. From an MDL lens [Rissanen, 1978, 1989], tokens in the middle still carry compressible regularity yet are not already compressed; they yield the largest expected reduction in description length per gradient unit.

What happens to the distribution of text? Banding changes *which* candidate spans survive screening, not their labels. Empirically, three shifts occur: *(i)* the empirical entropy histogram loses its "fat tails" and concentrates in a gentle dome—evidence that rote and junk are filtered; *(ii)* span–entropy heatmaps show fewer contiguous cold (rote) or hot (garbage) segments; *(iii)* the accepted set's type–token curve regains curvature (Heaps–b stabilizes or rises), indicating that novelty enters in *learnable* form. These shifts are precisely what downstream diagnostics (T2/T3) need to preserve heterozygosity across domains and time.

Planned visuals for T1 (drop–in figure set). *F1 (Entropy dome).* Two small panels per window: *(left)* kernel density of H_{seq} for **candidates** with the band $\mathcal{B}$ shaded; *(right)* the same for **accepted**. The band visibly trims the left/right tails while leaving a clean mid–mass. Caption highlights target band–pass (θ_H) and median entropy shift.

F2 (Span–entropy carpet). For 8 randomly sampled accepted sequences, plot a token–time heatmap of $H_{\text{span}}(s;\ell)$ (row = sequence, column = span start). Mark violations that would have failed the local left/right span checks. Caption: "*Banding prunes rote pockets and junk bursts while retaining structured uncertainty.*"

F3 (Band drift & stability). Line plots across R1→R4: band center and width; band–pass %; ECE on the accepted set. The desired pattern: stable band–pass near θ_H, non–increasing ECE, and a non–shrinking width (guarded by $\Delta_{\min}$). Caption calls out any adaptive nudges taken.

F4 (Ablation radar). Radar chart comparing *No Band* vs *Seq–band* vs *Joint–band* across five axes: repetition, Heaps–b, CVaR–NLL (hard slice), ECE, acceptance rate. Caption: "*Joint band best controls repetition and calibration without starving novelty.*"

Fisher information view (why mid–entropy is data–efficient). Token–level Fisher information is $I(\theta) = \mathbb{E}[\nabla \log p_\theta(y|x) \nabla \log p_\theta(y|x)^\top]$. In the

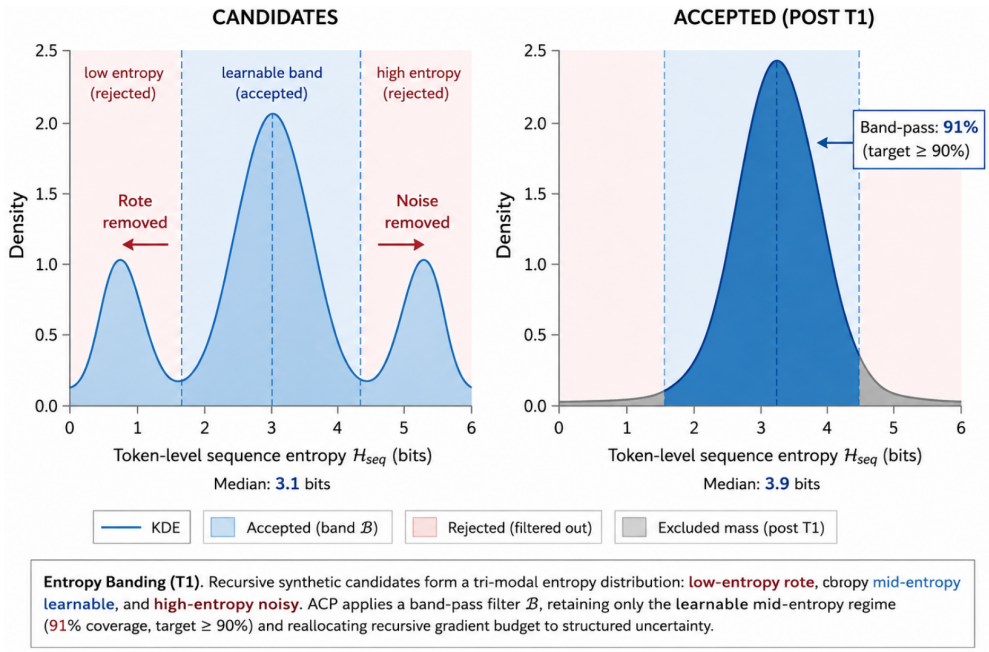


Figure 11: **Entropy Dome (Candidates vs. Accepted).** Two side-by-side KDE panels show token-level sequence entropy H_{seq} over the candidate pool and the post-T1 accepted pool. The shared shaded band $\mathcal{B} = [\underline{h}, \overline{h}]$ marks the admissible mid-entropy region. Candidates exhibit mild low-/high-entropy tails corresponding to rote head templates and noisy synthetic junk, while accepted samples concentrate into a centered "dome" with high band-pass rate, e.g., 91% against a target $\geq 90\%$. T1 therefore reallocates recursive gradient budget away from trivially predictable or degenerate spans and toward structured, learnable uncertainty.

left basin, $p_\theta(y|x)$ saturates and the score $\nabla \log p_\theta$ vanishes—Fisher collapses. In the right basin, p_θ is near–uniform and the score is high–variance but *uninformative* about structure (directions cancel). The middle basin yields *peaked but non–saturated* posteriors: the score has magnitude and directional coherence, concentrating Fisher along *task–relevant* subspaces. This explains why, under fixed gradient budget, middle–band tokens produce larger, stabler generalization gains than either extreme.

Calibration and the middle band. Overconfident mispredictions concentrate in the left basin (the model says "certain" but is wrong); underconfident correctness dominates the right basin. Banding reduces both: it removes the pathological left where ECE spikes *and* the chaotic right where Brier loss inflates. Reporting ECE/Brier on *accepted* items thus becomes a meaningful safety check: if ECE rises despite banding, the accepted set is drifting stylistically (often a T4 issue) or losing tails (T2 issue).

How to pick the band without hand–tuning. We recommend setting the initial middle as the $[15\%, 85\%]$ probe quantiles and then *learning* two hyper–knobs online: *(a)* a *center–nudge* that shifts $\mathcal{B}$ if ECE rises while repetition stays low (move slightly up in entropy), and *(b)* a *width–nudge* that widens $\mathcal{B}$ if Heaps–b dips but calibration is stable (add higher–entropy room for innovation). Both nudges are bounded (max 0.1 bits/token per window) and are only applied after checking T2/T3 metrics to avoid compensating for missing tails with noise.

Micro–failure modes to watch (and what they look like). *Template seepage.* Banding passes sequences with mid–entropy overall, but hidden low–entropy pockets (boilerplate) recur. *Signature:* span–left violations cluster at fixed offsets; repetition increases without Heaps–b gain. *Fix:* tighten span–min entropy and T2 template quotas; rotate a style critic (T4).

Noisy montage. Mid–entropy comes from stitching unrelated snippets, not coherent uncertainty. *Signature:* span–right violations spike, ECE improves little. *Fix:* introduce a *coherence critic* (embedding–based storyline smoothness) in T4; slightly lower $\overline{h}$.

Band chasing the head. As the student improves, the probe's quantiles slide left, moving $\mathcal{B}$ with them. *Signature:* band width shrinks; repetition

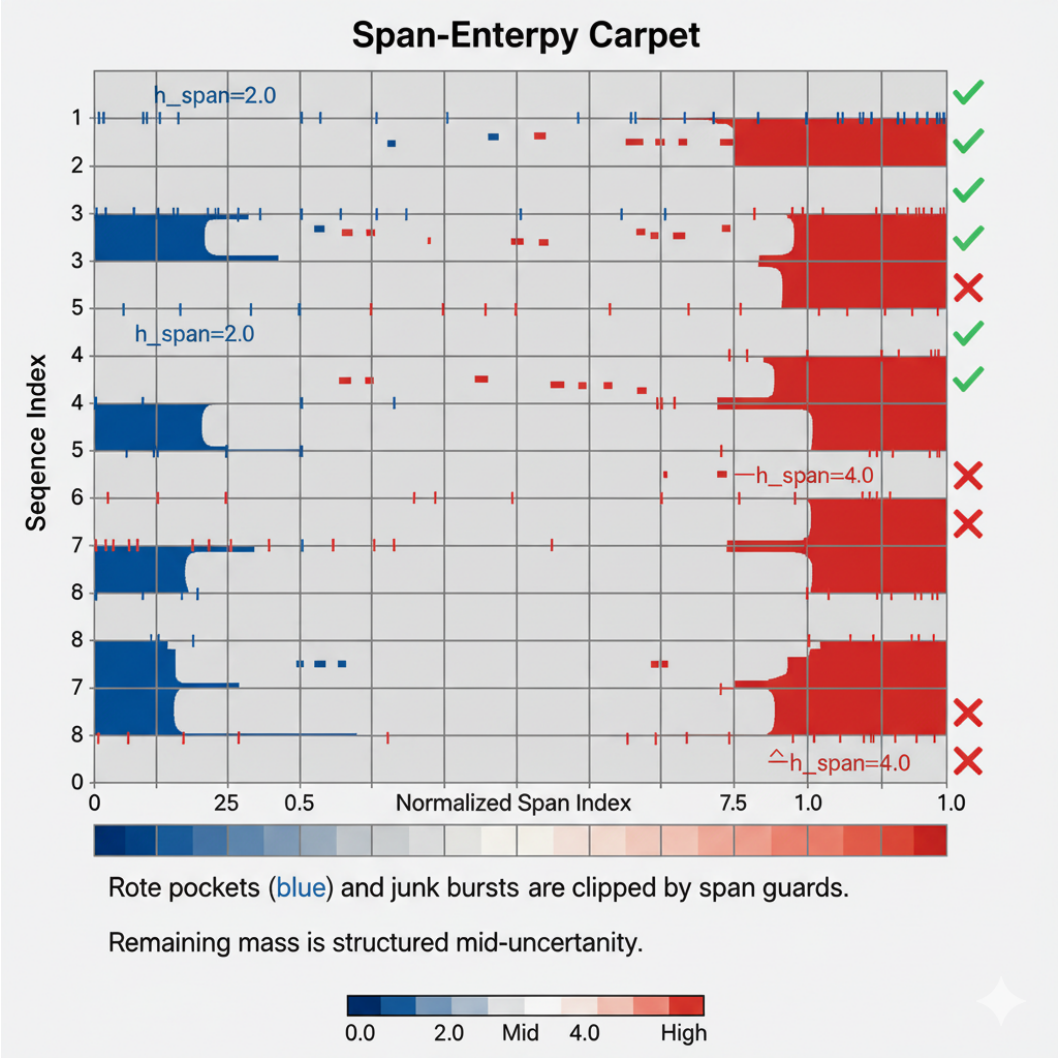


Figure 12: **Span-Entropy Carpet (Local Pockets of Rote/Junk).** Sliding-window span entropy $H_{\text{span}}(s;\ell)$, with $\ell = 64$, exposes local failures that sequence-level entropy can hide. Blue ticks mark low-entropy rote spans below $\underline{h}_{\text{span}}$; red ticks mark high-entropy bursts above $\overline{h}_{\text{span}}$. Per-row checks/crosses indicate whether the sequence passes span guards. The carpet shows how T1 removes hidden boilerplate and incoherent montages inside otherwise acceptable sequences, preserving *structured* uncertainty for recursive training.

rises slowly; tails stagnate. *Fix:* freeze center drift, widen toward high side, and enforce T2 tail floors.

Cross–lingual and domain specifics. Entropy distributions differ by morphology and tokenization. For agglutinative scripts, span entropy is naturally higher; for code/math, low entropy doesn't always imply rote—*but* template duplication is more dangerous. We therefore estimate $\mathcal{B}$ *per language/domain* and enforce *relative* rather than absolute thresholds (quantiles, not fixed bits). Diagnostics—especially Heaps–b and repetition—are also reported per domain to ensure the middle band isn't silently starving low–resource languages.

Interplay with counterfactual recompositions. Counterfactuals can artificially inflate entropy (e.g., unit or locale flips unfamiliar to the student) while preserving label semantics. Banding before *and* after the transform ensures that we accept *learnable* counterfactuals: those that increase support without crossing into the noise basin. In practice, we keep a small quota of "challenge" counterfactuals slightly above the upper band edge (e.g., the 90–92% region) to encourage exploration, but require T2 novelty floors and T4 critic pass–rates for admission.

An intuitive toy: mixture of experts on a ring. Suppose tokens are generated by latent "motifs" $\phi \in [0, 2\pi)$ on a ring; the student starts with confidence spikes around a few angles (head modes) and near–uniform elsewhere.

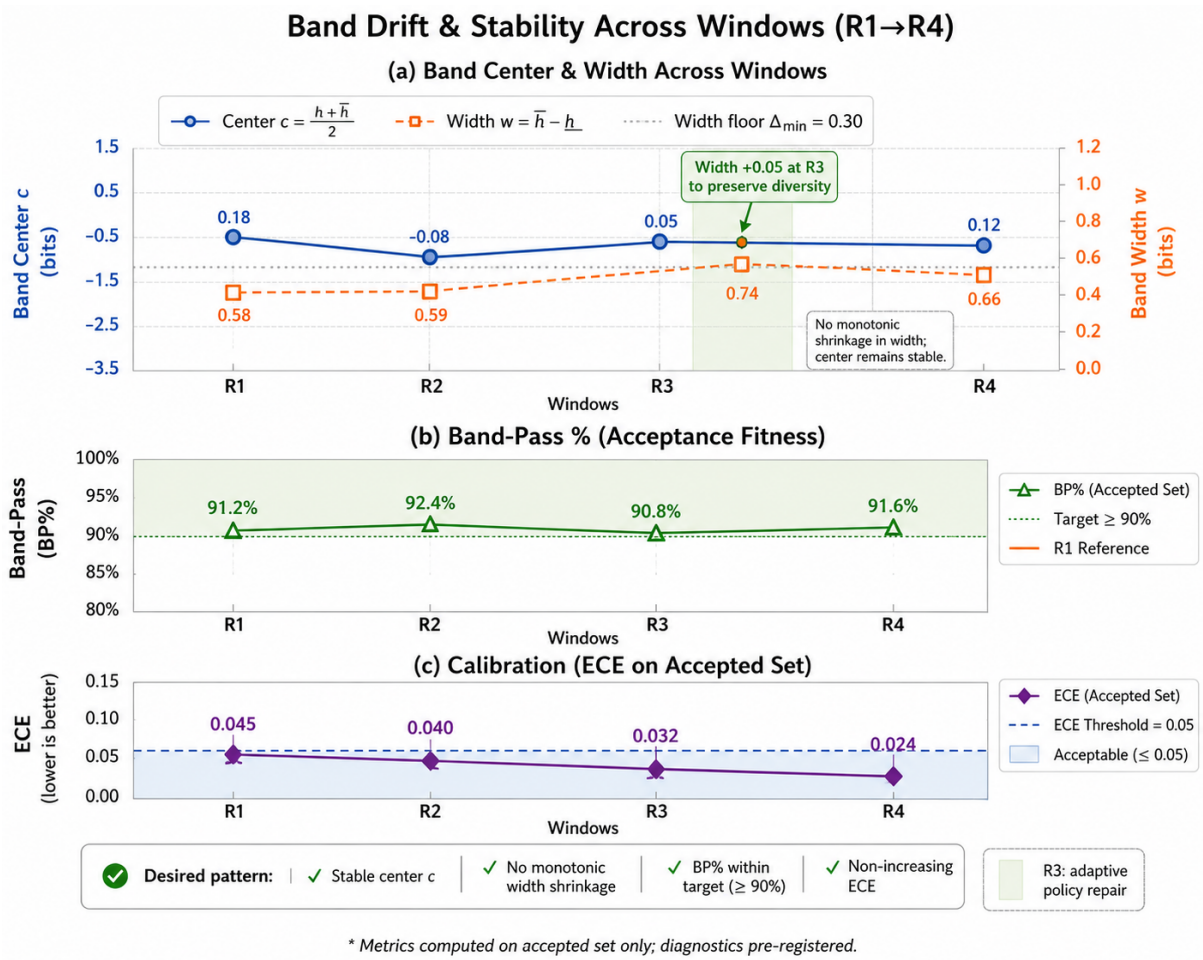


Figure 13: **Band Drift & Stability Across Windows** ($R1 \rightarrow R4$)**.** Three synchronized panels audit T1 across recursive windows: *top*: band center $c = \frac{\underline{h} + \overline{h}}{2}$ and width $w = \overline{h} - \underline{h}$, with width floor $\Delta_{\min}$; *middle*: band-pass percentage $\mathrm{BP}\%$ with target stripe, e.g., $\geq 90\%$; and *bottom*: expected calibration error (ECE) on the accepted set. A healthy run shows stable center, no monotonic width shrinkage, $\mathrm{BP}\%$ within target, and non-increasing ECE. Annotated policy repairs, such as width expansion at $R3$, make the screening rule auditable rather than post-hoc.

Left–basin updates sharpen the spikes (mass concentrates); right–basin updates wobble aimlessly (phase–incoherent). Middle–band updates—those near the flanks of each spike—*move the spike's shoulders outward*, effectively widening coverage and encouraging new modes to emerge. This picture mirrors what we observe: banding pushes probability mass *sideways* into under–covered neighborhoods rather than *upwards* at already–covered peaks.

Diagnostics & tripwires (reported per window). *(D1) Band–pass %* and *median* H_{seq} on accepted vs. rejected; *(D2) calibration* on accepted items (ECE, Brier) and a held–out hard slice [Guo et al., 2017]; *(D3) span–level violation rates* for left/right tails. **Tripwires:** any *two consecutive windows* with (i) band–pass $< \theta_H$, (ii) ECE $\uparrow$ on the banded set, or (iii) asymmetric span–violations (e.g., left–tail spikes) triggers adaptation: widen/nudge the band, rotate critics that penalize sameness (T4), and/or tighten T2 floors.

Ablations and expected outcomes. *No Band:* repetition/burstiness $\uparrow$, Heaps–b $\downarrow$, tail QA $\downarrow$, ECE $\uparrow$. *Seq–only vs. Joint band:* the joint $(H_{\text{seq}}, \overline{\text{sp}})$ band rejects sequences with pathological pockets (low–entropy islands, heavy–tailed surprisal) without starving novelty. + *Span min–entropy:* removes template boilerplate embedded in mid–entropy sequences. *Probe choice:* on–policy probes are adaptive; fixed references are conservative early on.

Implementation blueprint (reproducible defaults). *Probes:* 50k sentences per window (seed

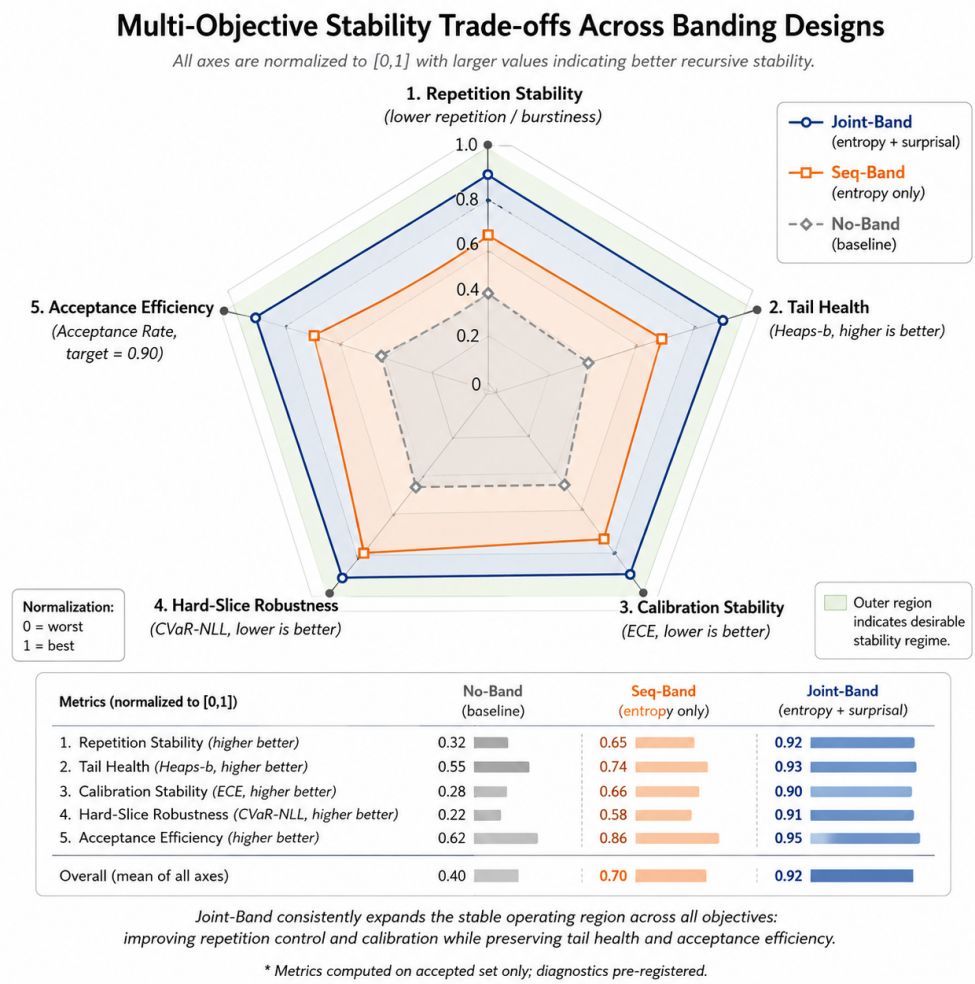


Figure 14: **Ablation Radar (No-Band vs. Seq-Band vs. Joint-Band).** Five normalized spokes compare entropy-band design choices: *Repetition/Burstiness* (inverted; higher is better), *Heaps-b* (higher is better), *CVaR-NLL on hard slices* (inverted), *ECE* (inverted), and *Acceptance Rate* scaled so the target equals 1. The *Joint-Band* variant improves repetition and calibration without sacrificing novelty, while *Seq-Band* improves stability relative to *No-Band*. This isolates the contribution of joint entropy–surprisal banding to tail health, hard-slice risk, and calibration.

human slice + small synthetic calibration). *Quantiles:* α=0.15 so $\mathcal{B}$ is $[15\%, 85\%]$; width floor $\Delta_{\min}$=0.3 bits/token. *Spans:* ℓ=64 tokens; span thresholds from probe quantiles. *Thresholds:* θ_H=0.9, $\rho_{\min}$=0.05, $\rho_{\max}$=0.05. *Joint band (optional):* highest–density contour covering 70% mass in $(H_{\text{seq}}, \overline{\text{sp}})$. *Reporting:* per–window violin plots of H_{seq} (accepted vs. rejected), ECE/Brier on accepted, span–violation histograms.

Planned visuals for T1 (to be placed as F1–F4). *F1—Entropy Dome.* KDE of H_{seq} for **candidates** and **accepted**, with $\mathcal{B}$ shaded; annotate band–pass % and median shift. *F2—Span–Entropy Carpet.* Token–time heatmaps of H_{span} for 8 accepted samples; mark left/right span violations. *F3—Band Drift and Stability.* Lines across R1→R4 for band center/width, band–pass %, and ECE on accepted. *F4—Ablation Radar.* No–Band vs Seq–Band vs Joint–Band across {repetition, Heaps–b, CVaR–NLL (hard slice), ECE, acceptance rate}.

Audit checklist (reviewer–reproducible). *(1)* Probe composition and entropy histograms; *(2)* $\underline{h}, \overline{h}$, target θ_H, realized band–pass per window; *(3)* ECE/Brier on accepted and on the hard slice; *(4)* span–violation rates with examples; *(5)* any band adjustments (center/width) with rationale tied to T2/T3/T4; *(6)* Heaps–b and repetition/burstiness for accepted vs. rejected. These fields render T1 *transparent, falsifiable, and auditable*.

Interlock with T2/T3/T4 (where T1 ends). Entropy Banding is a *band–pass filter on learnability*.

It does *not* by itself guarantee rare–token survival (T2), prevent style/recency bubbles (T3), or maintain multiple reasoning routes (T4). But without T1, T2–T4 must fight head–rote and junk simultaneously, wasting statistical power. In ACP, T1 clears the runway; T2–T4 then enforce tail/novelty throughput, coverage balance, and mechanism diversity—together preserving **heterozygosity** and preventing collapse.

A.3 Technique 2: Tail–Mass & Novelty Floors with Strict Near–Duplicate Caps

Objective and rationale. Collapse shows up first in the *tails*: rare tokens/phrases vanish, Heaps' growth flattens, and type–token diversity gives way to paraphrastic clones. **T2** establishes three coupled controls that are *provenance–agnostic* yet precise: *(i)* enforce a **Zipf–tail mass floor** so a fixed fraction of update budget flows to low–frequency types; *(ii)* enforce a **novelty floor** so each window contributes genuinely *out–of–support* n–grams (not just surface rephrasings); and *(iii)* impose **strict near–duplicate caps** in lexical, semantic, and template spaces to block clone inflation. Together these keep the token ecology *heterozygous*: rare alleles survive, novel alleles arrive, and clonal expansions are checked.

Preliminaries and sets. Let $\mathcal{V}$ be the vocabulary, $r(t)$ the Zipf rank of type t by frequency in the *accepted* pool up to window $w{-}1$, and fix a tail quantile $q \in (0,1)$ (e.g., $q{=}0.1$). Define the *tail set*

$$\mathcal{T}_q \;=\; \{t \in \mathcal{V} :\; r(t) > \lfloor (1{-}q)\,|\mathcal{V}| \rfloor\}.$$

Let G_n be the set of all observed n–grams up to $w{-}1$. For a candidate batch B, denote by $p_B(t)$ the empirical token frequency of type t inside B, by $U_n(B)$ the set of *new* n–grams in B (i.e., $U_n(B) = \{g \in G_n(B) \setminus G_n\}$), and by $\nu(B)$ a semantic *template* signature (defined below).

(T2.a) Zipf–tail mass floor. Define the tail–mass of B as

$$M_{\text{tail}}(B) \;=\; \sum_{t \in \mathcal{T}_q} p_B(t).$$

We **require** $M_{\text{tail}}(B) \geq \eta_q$ with η_q a fixed floor (e.g., 8% for $q{=}0.1$), *and* we allocate a minimum fraction of batch *spans* whose *median* token lies in $\mathcal{T}_q$. This couples token–level and span–level coverage so tails are not satisfied by isolated singletons. To avoid adversarial inflation via tail spam, we also cap per–type burstiness:

$$\max_{t \in \mathcal{T}_q} \frac{\text{count}_B(t)}{|B|} \;\leq\; \beta_{\text{tail}} \quad (\text{e.g., } \beta_{\text{tail}} = 0.5\%).$$

These two constraints guarantee *throughput* of diverse tail tokens without allowing any one rare type to dominate.

(T2.b) Novelty floor (out–of–support n–grams). For $n \in \{3,4,5\}$, compute the *unique novelty rate*

$$\phi_n(B) \;=\; \frac{|U_n(B)|}{\text{total } n\text{–grams in } B}.$$

We **require** $\phi_n(B) \geq \gamma_n$ (e.g., $\gamma_3{=}5\%$, $\gamma_4{=}3\%$, $\gamma_5{=}1\%$), with a *window–level* floor on the cumulative additions $|U_n(\text{window})|$. To prevent novelty from being *noise masquerading as new*, we couple ϕ_n to T1's learnability: a candidate only contributes to U_n if its sequence and span entropies pass banding; thus "new" means *learnably out–of–support*. We also monitor the *Good–Turing* mass of unseen types to ensure the growing frontier is statistically meaningful.

(T2.c) Strict near–duplicate caps (lexical, semantic, template). We maintain three *similarity graphs* over the candidate pool $\mathcal{C}$:

Lexical graph G^{lex}: edges connect documents with Jaccard $J \geq \tau_{\text{lex}}$ on k–shingles (e.g., $k{=}5$, $\tau_{\text{lex}}{=}0.85$) found via LSH/MinHash.

Semantic graph G^{sem}: edges connect with cosine similarity $\sigma \geq \tau_{\text{sem}}$ in a sentence–embedding space (e.g., SBERT), with adversarially robust pooling to avoid exact–copy laundering.

Template graph G^{tpl}: edges connect equal *templates* $\nu(\cdot)$ derived from delexicalized patterns (slotting numbers, entities, units, locales) so surface swaps do not create fake diversity.

For each graph, we cap connected–component sizes: for any component C, require $|C| \leq \kappa$ and per–batch degree $\deg_B(x) \leq d_{\max}$ (e.g., $\kappa{=}3$, $d_{\max}{=}1$). We also apply a *global* window cap on the share of accepted items that are within ϵ of a previous accepted item under *any* metric, preventing slow clone accretion across windows.

Why this triad prevents tail death. Zipf floors route gradient to the long tail. Novelty floors ensure the tail *replenishes* with unseen compositional

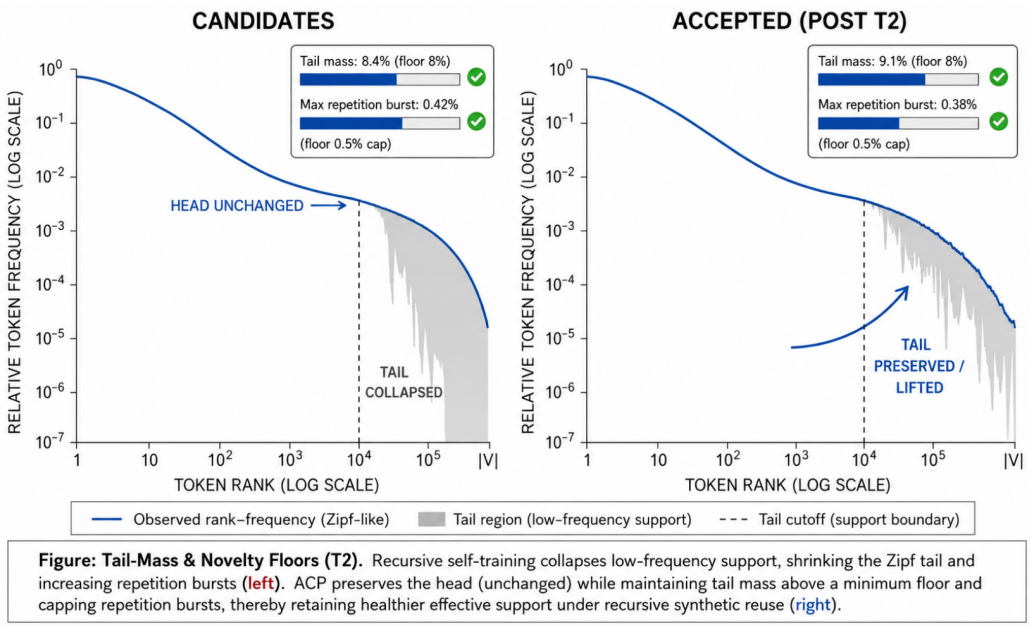


Figure 15: **Zipf Tail–Mass.** Log–log rank–frequency curves compare *candidates* and *accepted* samples. The shaded last-10% ranks mark the Zipf tail. Insets report tail mass against the floor η_q and maximum tail burstiness against the cap β_{tail}. Accepted samples preserve or lift tail density while respecting burstiness caps, showing that T2 keeps low-frequency support alive under recursive reuse.

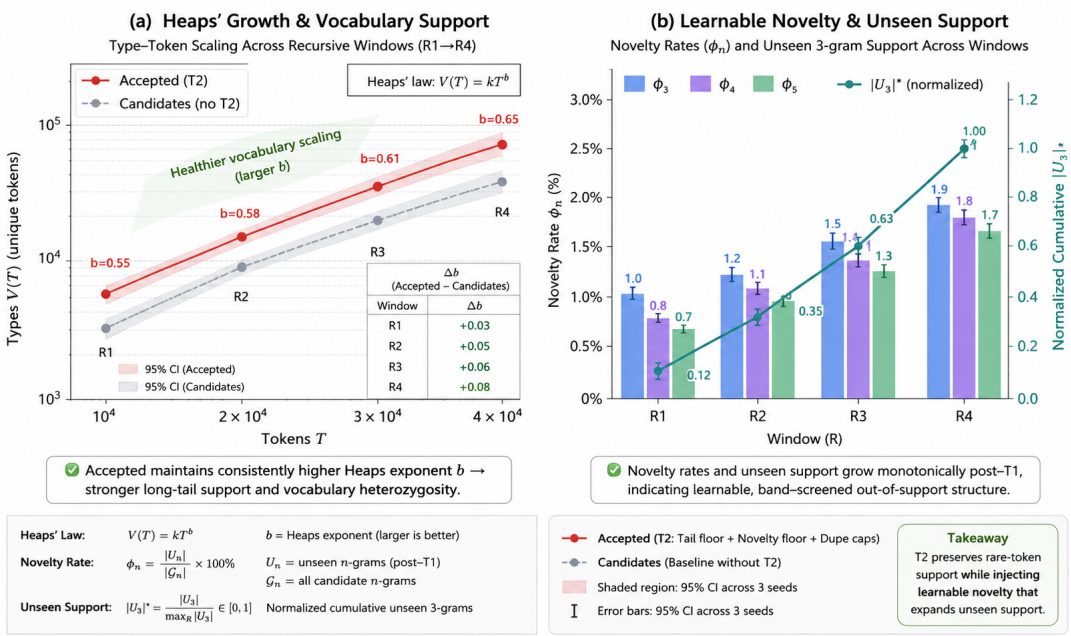


Figure 16: **Heaps' Growth & Learnable Novelty.** Left: type–token Heaps curves across $R1 \to R4$, with fitted exponent b. Accepted curves rise or stabilize above candidate curves, indicating healthier vocabulary growth. Right: novelty rates ϕ_n for $n \in \{3, 4, 5\}$, cumulative unseen n-gram counts $|U_n|$, and Good–Turing unseen mass. Since novelty is measured after T1 screening, the accepted samples inject *learnable* rather than noisy out-of-support structure.

structures rather than permutations of the head. Duplicate caps stop *frequency hijacks* whereby semantically identical clones crowd out rare types while still passing naive novelty counters. The triad therefore *widens* effective support (more types live), *deepens* support (more distinct n–grams), and *scrambles* the over–represented head templates, keeping heterozygosity high.

Interaction with T1/T3/T4. T1 filters to *learnable* candidates so U_n is not polluted by degener-

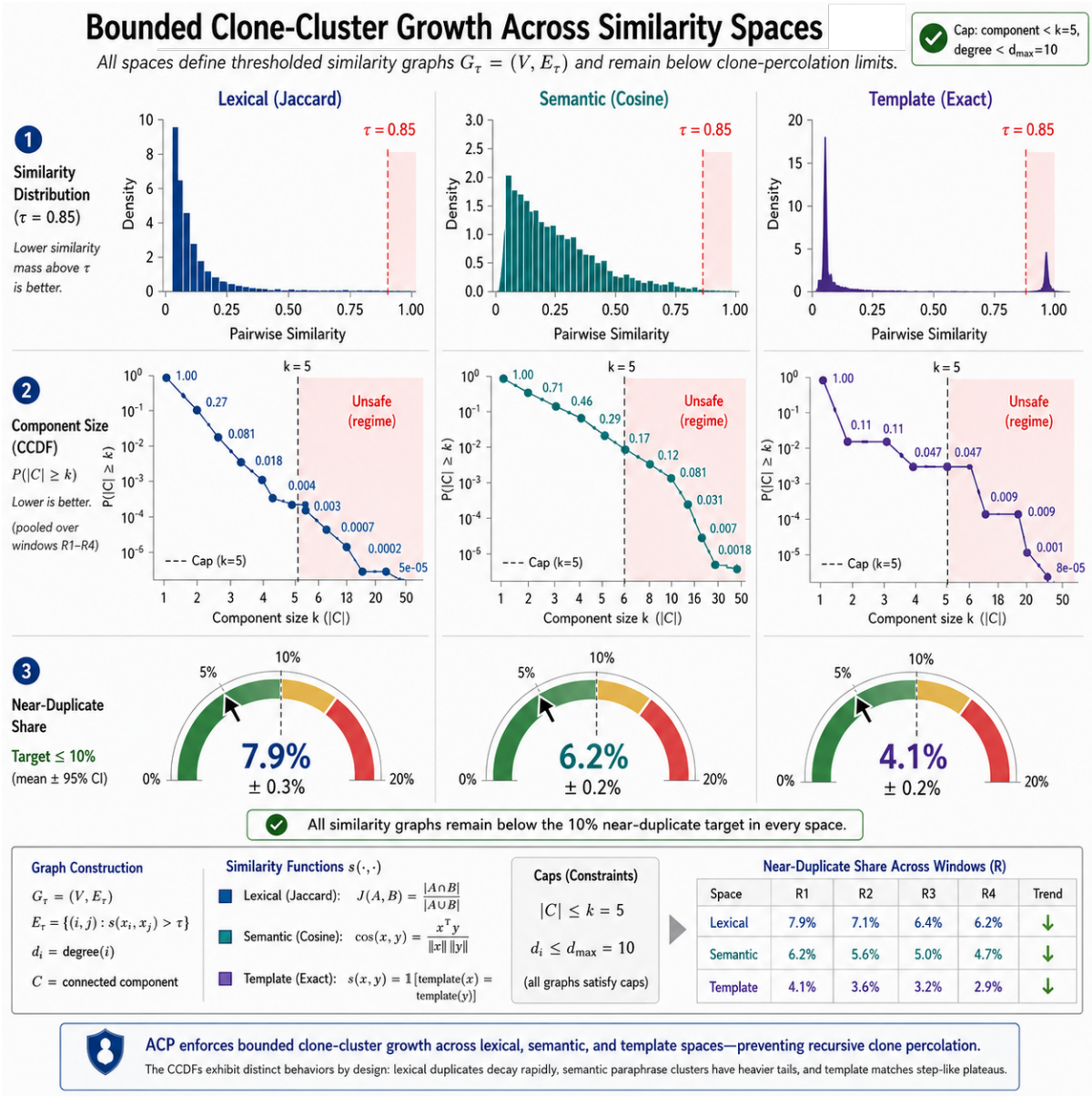
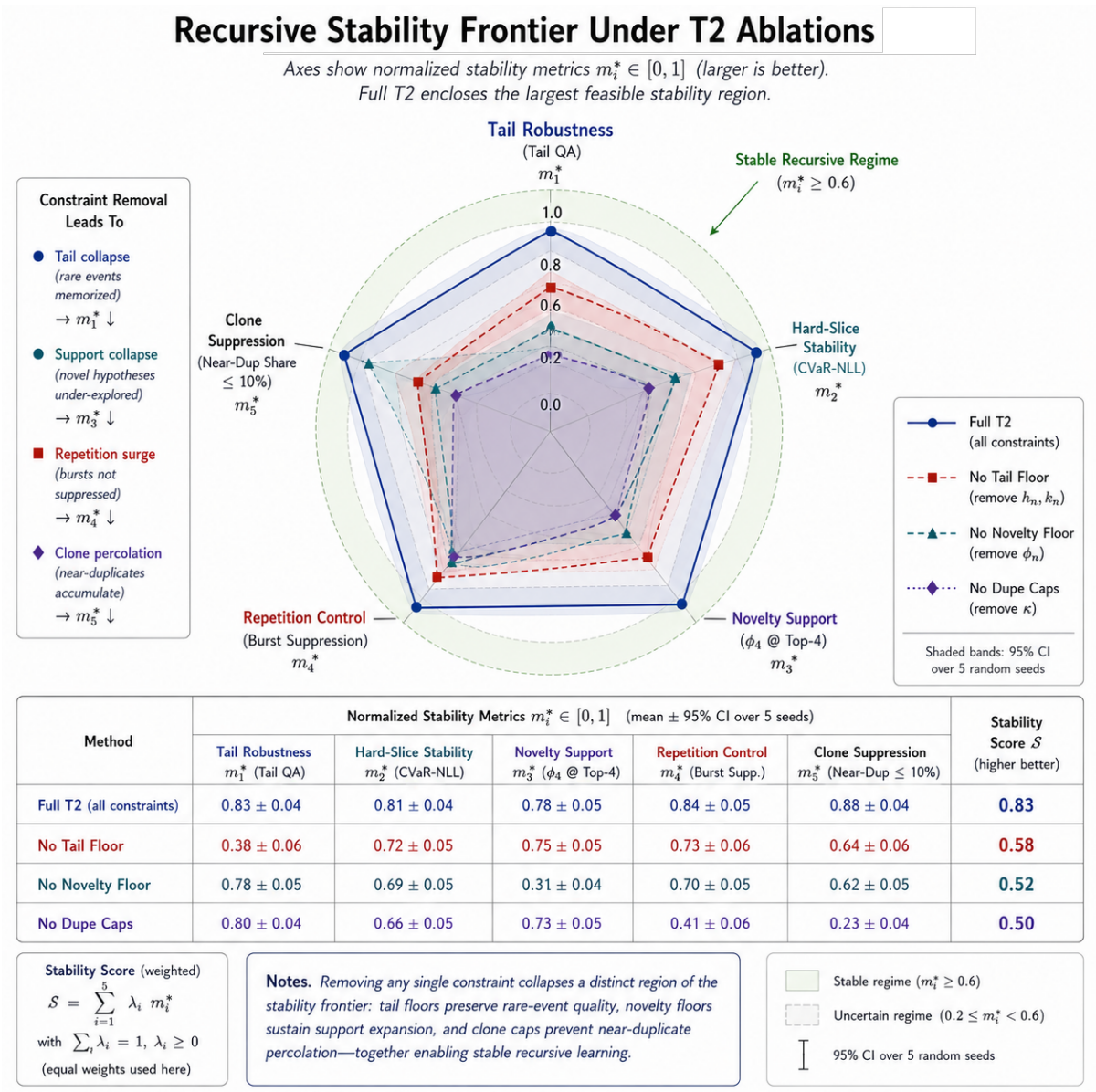


Figure 17: **Near–Duplicate Control.** T2 suppresses paraphrastic clone inflation across three similarity spaces: lexical Jaccard, semantic cosine, and delexicalized-template matching. Top panels show similarity histograms with thresholds τ; middle panels show connected-component sizes with cap κ; and bottom summaries report near-duplicate share against the target $\leq 10\%$. The caps remain satisfied across spaces, preventing apparent diversity from being inflated by near-clones.

ate strings. T3 spreads tail/novelty mass across *domains and time*, stopping single–style bubbles (e.g., tail satisfied only by code or only by recent news). T4 forces *mechanism diversity*: multiple reasoning routes generate different n–gram frontiers for the same task, increasing the chance that U_n carries *structured* novelty rather than noisy detours.

Diagnostics and tripwires (per window). *(D1) Tail mass:* realized M_{tail} vs. floor η_q; histogram of tail–token shares per batch; burstiness maxima vs. β_{tail}. *(D2) Type–token growth:* Heaps' exponent b on accepted vs. rejected; differential Δb across windows. *(D3) Novelty:* ϕ_n (for n=3, 4, 5), cumulative $|U_n|$, Good–Turing unseen mass; proportion of *learnable* novelties (those also passing T1). *(D4) Clone control:* size distribution of connected components in $G^{\text{lex}}, G^{\text{sem}}, G^{\text{tpl}}$; fraction of items within ϵ of prior acceptances. *(D5) Downstream tail health:* accuracy and CVaR–NLL on tail–focused eval slices; rare–entity QA and long–tail factual probes.

Tripwires: trigger *Adapt* if any occur for two



Method	Normalized Stability Metrics $m_i^* \in [0,1]$ (mean ± 95% CI over 5 seeds)					Stability Score $\mathcal{S}$ (higher better)
	Tail Robustness m_1^* (Tail QA)	Hard-Slice Stability m_2^* (CVaR-NLL)	Novelty Support m_3^* (ϕ_4 @ Top-4)	Repetition Control m_4^* (Burst Supp.)	Clone Suppression m_5^* (Near-Dup ≤ 10%)	
Full T2 (all constraints)	0.83 ± 0.04	0.81 ± 0.04	0.78 ± 0.05	0.84 ± 0.05	0.88 ± 0.04	**0.83**
No Tail Floor	0.38 ± 0.06	0.72 ± 0.05	0.75 ± 0.05	0.73 ± 0.06	0.64 ± 0.06	**0.58**
No Novelty Floor	0.78 ± 0.05	0.69 ± 0.05	0.31 ± 0.04	0.70 ± 0.05	0.62 ± 0.05	**0.52**
No Dupe Caps	0.80 ± 0.04	0.66 ± 0.05	0.73 ± 0.05	0.41 ± 0.06	0.23 ± 0.04	**0.50**

Stability Score (weighted)

$$\mathcal{S} = \sum_{i=1}^{5} \lambda_i \, m_i^*$$

with $\sum_i \lambda_i = 1$, $\lambda_i \geq 0$

(equal weights used here)

Notes. *Removing any single constraint collapses a distinct region of the stability frontier: tail floors preserve rare-event quality, novelty floors sustain support expansion, and clone caps prevent near-duplicate percolation—together enabling stable recursive learning.*

Stable regime ($m_i^* \geq 0.6$)

Uncertain regime ($0.2 \leq m_i^* < 0.6$)

95% CI over 5 random seeds

Figure 18: **T2 Ablation Radar.** Full T2 is compared against ablations {*No Tail Floor*, *No Novelty Floor*, *No Dupe Caps*} across six normalized axes: Tail QA, CVaR-NLL ↓, Heaps-b, ϕ_4, Repetition ↓, and Near-duplicate share ↓. All axes are oriented so larger values indicate better recursive stability. Full T2 expands the stable operating region by preserving tail quality and novelty while minimizing repetition and clone growth.

consecutive windows—(i) $M_{\text{tail}} < \eta_q$, (ii) $\phi_n < \gamma_n$ for any n, (iii) b decreases while clone component sizes grow, (iv) tail–slice CVaR–NLL worsens, (v) duplicate–component caps are repeatedly saturated. Remediation: raise η_q or γ_n; tighten $\tau_{\text{lex}}, \tau_{\text{sem}}, \tau_{\text{tpl}}$ or lower κ; reweight to under–covered domains (handed off to T3); rotate critics to penalize paraphrastic sameness (T4).

Failure modes and mitigations. *Tail spam.* Inflating M_{tail} with meaningless rare tokens. *Mitigate:* require *span–level* tail presence and T1 pass; couple to task loss on tail slices. *Template laundering.* Many surface variants of one template pass novelty. *Mitigate:* enforce *template* duplicate caps via delexicalization; measure unique templates per window. *Semantic echo.* High embedding similarity across putatively novel items. *Mitigate:* lower τ_{sem}, increase $d_{\max}$ penalties, and add *route–divergence* auditing (T4).

Parameterization (reproducible defaults). q=0.1, η_q=8%, β_{tail}=0.5%; novelty floors γ_3=5%, γ_4=3%, γ_5=1%; MinHash k=5 shingles, Jaccard threshold τ_{lex}=0.85; semantic cosine threshold τ_{sem}=0.85; template exact–match

via delexicalization rules with unit/locale/NER slotting; caps κ=3, $d_{\max}$=1; window–level max share of near–dups across any metric $\leq 10\%$.

Ablations we expect to see. *No tail floor:* $M_{\text{tail}} \downarrow$, Heaps–$b \downarrow$, rare–slice CVaR–NLL $\uparrow$, calibration drifts. *No novelty floor:* $\phi_n \downarrow$, $|U_n|\downarrow$, Good–Turing unseen mass stagnates; performance plateaus despite volume. *No dupe caps:* component sizes swell, repetition/burstiness $\uparrow$, apparent novelty rises but template diversity *falls*; long–tail QA drops.

Intuition: a population–genetics picture. Think of tokens/types as alleles. Zipf floors enforce *viable effective population size* in the tail; novelty floors inject *mutation* at a controlled rate; duplicate caps limit *clonal expansions*. The trio sustains *heterozygosity* across windows—precisely the quantity that collapses under recursive self–feeding.

Planned visuals for T2 (F5–F8). *F5 — Zipf Tail–Mass Plot (Candidates vs. Accepted).* Log–log rank–frequency curves with the same tail quantile q highlighted. Accepted curve preserves or lifts the tail slope; inset shows M_{tail} vs. floor η_q and per–type burstiness maxima vs. β_{tail}.

F6 — Heaps' Growth & Novelty. Two panels: (left) type–token curves with fitted Heaps' exponent b for candidates and accepted (R1→R4 lines); (right) bar/line combo for ϕ_n (n=3, 4, 5) and cumulative $|U_n|$, with Good–Turing unseen mass as a dashed overlay.

F7 — Near–Duplicate Control. Three small multiples for lexical/semantic/template spaces: histograms of pairwise similarity with a vertical cut at τ; connected–component size distributions; a donut showing share of near–dups in the window (target $\leq 10\%$). Callouts indicate κ and $d_{\max}$ hits.

F8 — T2 Ablation Radar. Radar for {tail QA, CVaR–NLL (tail), Heaps–b, ϕ_4, repetition/burstiness, near–dup share} comparing *Full T2* vs. *No Tail Floor* vs. *No Novelty Floor* vs. *No Dupe Caps*. Full T2 should dominate on tail QA, Heaps–b, and low dup share.

Short captions (to accompany F5–F8). *F5:* "Accepted sets keep Zipf tails alive: tail mass M_{tail} meets floor and burstiness is capped." *F6:* "Heaps' b and out–of–support n–grams grow across windows; novelty is *learnable* (post T1)." *F7:* "Strict caps curb clones across lexical, semantic, and template views; near–dup share stays $\leq 10\%$." *F8:* "Removing any T2 leg harms tail QA and novelty while inflating repetition or clones—confirming causality."

A.4 Technique 3: Domain Isotropy with Temporal Stratification (preventing single–style/recency lock–in)

Objective and threat model. Synthetic loops readily *over–specialize*: a few **styles** (e.g., newswire, Q&A, code) or a recent **time window** dominate; the student calibrates to these micro–regimes, and generalization to under–represented styles/eras decays—classic *dataset shift* and *concept drift* phenomena [Quinonero-Candela et al., 2008; Gama et al., 2014]. **T3** imposes *isotropy* across domains and *stratification* across time so that (i) no single style becomes a gravitational well and (ii) no recent synthesis wave overwhelms earlier windows. The control is *provenance–agnostic*: it regulates the *geometry of coverage* rather than who authored the text.

Partitions and priors. Let $\mathcal{D} = \{1, \ldots, D\}$ index domains/styles (e.g., news, web, code, math, legal; or unsupervised clusters), and $\mathcal{W} = \{1, \ldots, R\}$ index synthesis windows (*R1→R4*). For any candidate or accepted pool $\mathcal{A}$, define empirical marginals

$$\hat{p}_{\mathcal{A}}(d) = \frac{N_{\mathcal{A}}(d)}{|\mathcal{A}|}, \qquad \hat{p}_{\mathcal{A}}(w) = \frac{N_{\mathcal{A}}(w)}{|\mathcal{A}|},$$

and the joint $\hat{p}_{\mathcal{A}}(d, w)$. Choose *smooth* target priors $\pi_D(d)$ and $\pi_W(w)$ (uniform or smoothed historical proportions), and a separable joint prior $\pi(d, w) = \pi_D(d)\pi_W(w)$ unless a deliberate domain–time coupling is desired.

Isotropy and stratification constraints. We enforce *upper bounds* on divergences between empirical coverage and priors:

$$\mathrm{KL}(\hat{p}_{\mathcal{A}}(d) \,\|\, \pi_D(d)) \leq \delta_D, \qquad \mathrm{KL}(\hat{p}_{\mathcal{A}}(w) \,\|\, \pi_W(w)) \leq \delta_W,$$

and *pairwise imbalance* caps

$$\max_d \frac{\hat{p}_{\mathcal{A}}(d)}{\pi_D(d)} \leq \rho_D, \qquad \max_w \frac{\hat{p}_{\mathcal{A}}(w)}{\pi_W(w)} \leq \rho_W.$$

To catch subtle coupling (e.g., a "news@R4" bubble), we also control the joint via total variation

$$\mathrm{TV}(\hat{p}_{\mathcal{A}}(d, w),\, \pi_D(d)\pi_W(w)) \leq \tau_{DW},$$

or, when w encodes a real temporal coordinate, a *Wasserstein–1* constraint along the time axis:

$$W_1(\hat{p}_{\mathcal{A}}(w),\, \pi_W(w)) \le \epsilon_W,$$

which penalizes *recency drift* by mass–transport distance rather than solely by ratios.

Acceptance test and repair sampler. Given a screened batch B (already passed **T1/T2**), compute provisional updates

$$\hat{p}'(d) = \frac{N_{\mathcal{A}}(d) + N_B(d)}{|\mathcal{A}|+|B|}, \qquad \hat{p}'(w) = \frac{N_{\mathcal{A}}(w) + N_B(w)}{|\mathcal{A}|+|B|},$$

and accept B *iff* all divergence/imbalance caps remain within bounds when moving from $\hat{p}$ to $\hat{p}'$. To avoid hard rejections when B is otherwise valuable (e.g., strong **T2** novelty), we apply a *repair sampler* that down–samples dominant (d, w) cells and up–samples under–covered cells inside B using importance weights

$$\omega(d,w) \;\propto\; \min\left\{\frac{\pi_D(d)}{\hat{p}_{\mathcal{A}}(d)}, \frac{\pi_W(w)}{\hat{p}_{\mathcal{A}}(w)}\right\}, \quad \text{with} \quad \omega \in [\omega_{\min}, \omega_{\max}],$$

then re–check the repaired $B^\star$ against the caps and **T1/T2** constraints.

Style isotropy beyond coarse domains. Hand labels are coarse and leaky. We therefore induce *style clusters* by embedding documents (sentence–transformers) and running k–means or spectral clustering.[2] Let $\mathcal{S} = \{1, \ldots, S\}$ be the clusters; we control *cluster entropy*

$$H_{\text{style}}(\mathcal{A}) \;=\; -\sum_{s=1}^{S} \hat{p}_{\mathcal{A}}(s) \log \hat{p}_{\mathcal{A}}(s)$$

with a floor $H_{\text{style}} \ge H_{\min}$ and a divergence cap $\mathrm{KL}(\hat{p}(s) \| \pi_S(s)) \le \delta_S$. This prevents "one–note" datasets even when domain labels look balanced.

Why isotropy+time caps prevent lock–in (mechanism sketch). Let θ be the student's parameters and ℓ the token–level NLL. Under domain mixture $d \sim \hat{p}(d)$, the expected gradient decomposes as

$$\nabla \mathbb{E}_d\, \mathbb{E}_{x|d}\, \ell(\theta; x) \;=\; \sum_d \hat{p}(d)\, \underbrace{\mathbb{E}_{x|d}\, \nabla \ell(\theta; x)}_{G_d}.$$

If $\hat{p}(d)$ is skewed, learning becomes *directionally biased* toward a few G_d directions; low–mass domains get insufficient corrective updates and calibration drifts [Guo et al., 2017]. Imposing $\mathrm{KL}(\hat{p}(d) \| \pi_D) \le \delta_D$ bounds the simplex distance to the target mixture, which in turn bounds the deviation of the actual gradient from the *desired* mixture gradient:

$$\Big\| \sum_d (\hat{p}(d) - \pi_D(d))\, G_d \Big\| \;\le\; \|\hat{p} - \pi_D\|_1 \max_d \|G_d\| \le \sqrt{2\,\delta_D} \max_d \|G_d\|,$$

by Pinsker's inequality. Hence limiting δ_D *uniformly* controls how far updates can rotate toward a single style. Analogously along time, if $\hat{p}(w)$ drifts to recent windows, the student overfits to *fresh* synthesis idiosyncrasies; the W_1 cap prevents mass collapse onto the last window by penalizing long transports.

Temporal stratification: rolling quotas and mild aging. We impose *rolling quotas* per time cell so each window contributes at least α_W of total tokens, and we age old windows by a mild *decay prior* $\tilde{\pi}_W(w) \;\propto\; \pi_W(w)\, \exp(-\lambda\, \mathrm{age}(w))$ with small λ to avoid fossilization while resisting recency spikes. **Tripwires** fire if: *(i)* any single w contributes $> \rho_W$ of tokens, or *(ii)* W_1 exceeds ϵ_W for two consecutive evaluation checkpoints.

Interplay with T1/T2/T4. **T1** ensures that rebalancing does not admit rote or junk from under–covered cells; **T2** guarantees that *within* each (d, w), tails and novelty are preserved; **T4** drives *mechanism diversity inside* every domain/time cell so isotropy does not devolve into uniform route uniformity.

Diagnostics & pre–registered tripwires. *(D1) Coverage:* $\mathrm{KL}(\hat{p}(d) \| \pi_D)$, $\mathrm{KL}(\hat{p}(w) \| \pi_W)$, $\mathrm{TV}(\hat{p}(d,w), \pi_D \pi_W)$, $W_1(\hat{p}(w), \pi_W)$; per–cell over–representation ratios $\hat{p}/\pi$.
(D2) Style entropy: H_{style} and cluster–dominance index $\max_s \hat{p}(s)$.
(D3) Performance balance: per–domain and per–time *CVaR–NLL* on hard slices [Rockafellar and Uryasev, 2000], and per–cell calibration (ECE).
(D4) Adapt log: sampler weights $\omega(d, w)$ and rejection/repair counts.
Tripwires (any holds twice consecutively): KL caps breached, W_1 rises, $H_{\text{style}} < H_{\min}$, or domain/time CVaR–NLL gaps widen; *Adapt* then (i) tightens $\delta_D, \delta_W, \tau_{DW}$, (ii) raises quotas for under–served cells, and (iii) rotates sameness critics toward the dominant style (**T4**).

[2] *Unsupervised style clusters* also capture register/format, e.g., *narrative*, *bullet points*, *code with comments*, *formal proofs*.

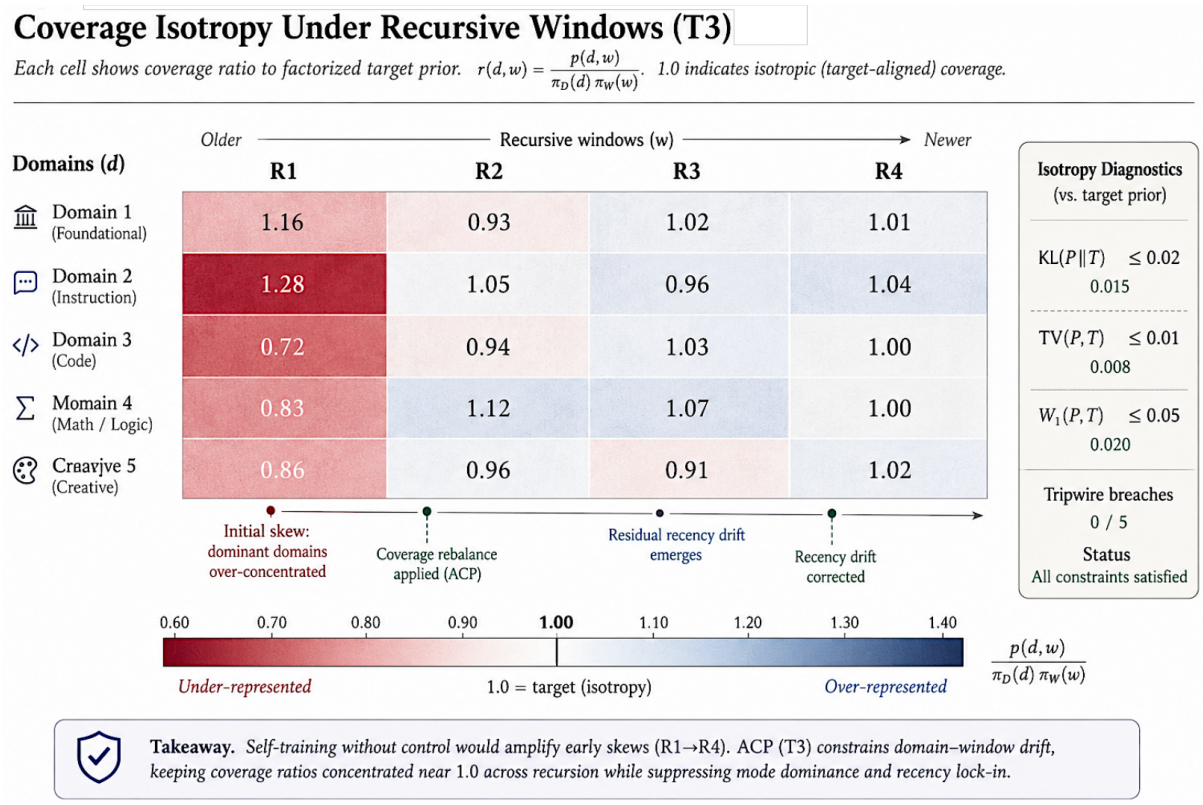


Figure 19: **Coverage Heatmap (Domain × Window).** Heatmap of coverage ratios $\hat{p}(d,w)/(\pi_D(d)\pi_W(w))$ with a diverging scale centered at 1.0. Out-of-tolerance cells are outlined. Margin bars report marginal coverage ratios $\hat{p}(d)/\pi_D(d)$ and $\hat{p}(w)/\pi_W(w)$ with cap ticks ρ_D and ρ_W. Header badges summarize KL_D, KL_W, TV_{DW}, and W_1, while repair-sampling annotations mark where T3 rebalances coverage.

Parameterization (transparent defaults). Priors: π_D uniform over D (or documented task–specific mix), π_W uniform across R1–R4. Caps: $\delta_D{=}\delta_W{=}0.02$ (Pinsker $\Rightarrow \|\hat{p}-\pi\|_1 \lesssim 0.2$), $\rho_D{=}\rho_W{=}2.0$ (at most $2\times$ over–representation), $\tau_{DW}{=}0.05$, ϵ_W equal to one time–bin width, $H_{\min}$ at the 70th percentile of H_{style} measured on the seed human slice. Quotas: $\alpha_W{=}15\%$ per window minimum. Repair weights: $\omega_{\min}{=}0.5$, $\omega_{\max}{=}2.0$.

Failure modes & mitigations. *Style bubble behind balanced domains.* Coarse "news" hides a dominant *template*. *Mitigate:* enforce style clusters and template–level caps; rotate a *template sameness* critic (**T4**).
Recency echo. Newest window subtly dominates due to prompt freshness. *Mitigate:* increase ϵ_W penalty, raise earlier–window quotas, inject *counterfactual recompositions* from older windows (**T4**).
Over–regularization. Tight caps starve valuable niche cells (e.g., legal proofs). *Mitigate:* allow *local* budget waivers when **T2** novelty and tail–slice CVaR–NLL show large gains; log exceptions.

Intuition: coverage as a spherical shell. Imagine each domain/time cell contributing a basis vector in a high–dimensional coverage simplex. Lock–in collapses the mass toward a few basis directions (a *spike*); isotropy spreads mass into a *shell* of roughly equal radius. The KL/TV/W_1 caps keep the empirical point near that shell; **T2/T4** then sculpt *what* lives on the shell (tails, novelty, mechanisms).

Ablations we expect to see. *No T3:* domain/time KL grows, W_1 drifts to the latest window, ECE degrades on under–served domains, and CVaR–NLL gaps widen; repetition falls only within the dominant style.
Caps only on domains (no time): cross–style metrics look stable but recency bias appears (older–era QA drops).
Time caps only: style bubbles form (cluster entropy drops) even as recency remains bounded.

Planned visuals for T3 (F9–F12).

- *F9 — Coverage Heatmap (Domain × Win-*

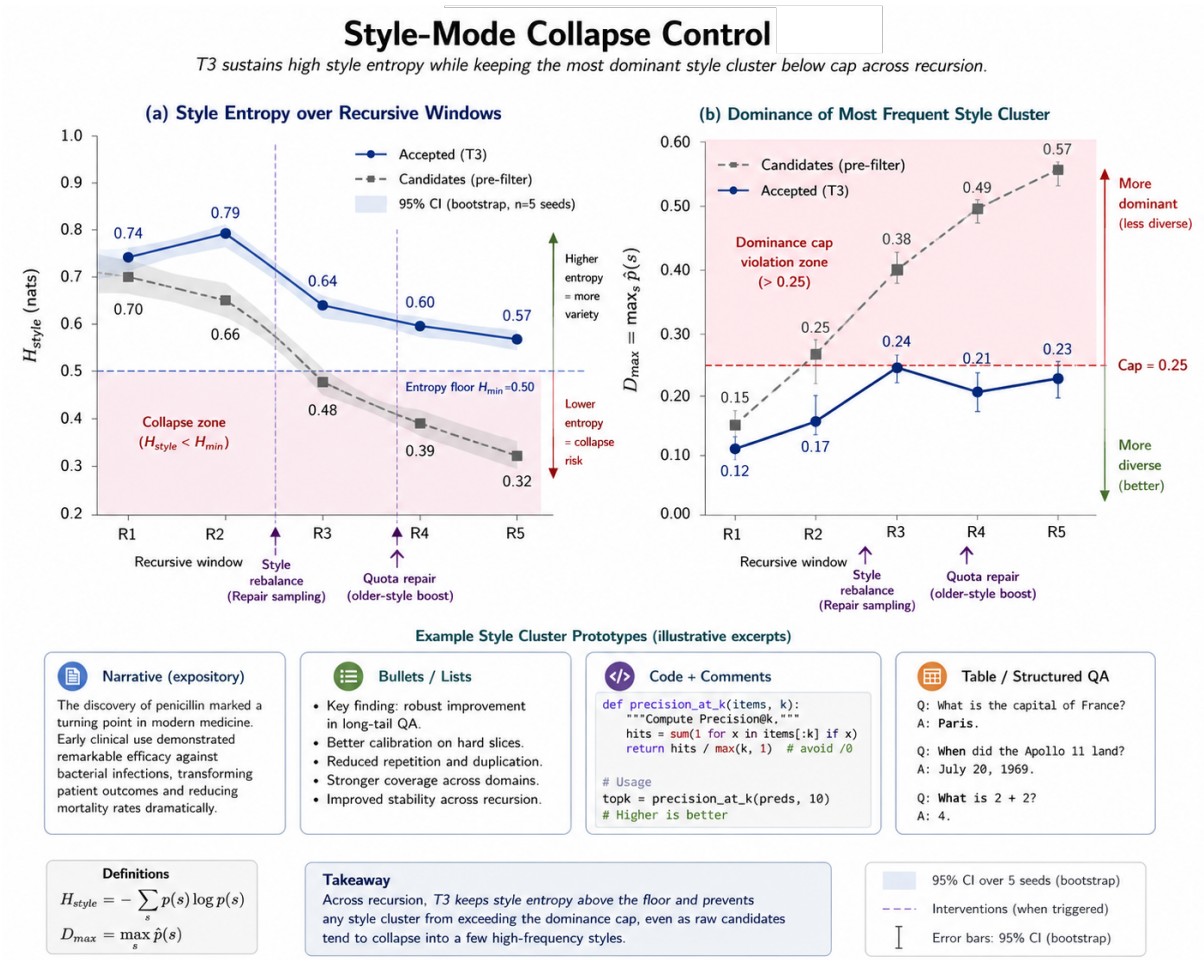


Figure 20: **Style Isotropy: Cluster Entropy & Dominance.** Left: style-cluster entropy H_{style} across recursive windows $R1 \rightarrow R4$, with the shaded floor band $H_{\min}$. Accepted samples remain above the entropy floor throughout recursion. Right: dominance index $\max_s \hat{p}(s)$ with cap line; exemplar tiles illustrate style variation across narrative, bullet, code, and structured formats. T3 prevents any single style cluster from dominating the accepted recursive pool.

dow). $D \times R$ matrix of $\hat{p}(d,w)/\pi(d,w)$; out–of–tolerance cells outlined; margins annotate $\mathrm{KL}(\hat{p}(d)\|\pi_D)$, $\mathrm{KL}(\hat{p}(w)\|\pi_W)$, and W_1.

- *F10 — Style Isotropy.* Two panels: (left) cluster entropy H_{style} over R1→R4 with a floor band; (right) dominance index $\max_s \hat{p}(s)$ with cap, plus exemplar snippets from top clusters to make "style bubble" visually obvious.
- *F11 — Recency Drift.* Temporal histograms of accepted tokens per window with W_1 trendline and quota bands; callouts where repair sampling was triggered.
- *F12 — T3 Ablation Radar.* Polygons for *Full T3*, *No Domain Caps*, *No Time Caps*, *No Style Clusters* across axes: domain/time KL (inverted), H_{style}, dominance index (inverted), per–domain CVaR–NLL (inverted), per–time ECE (inverted).

Short captions (for F9–F12). *F9:* "Coverage stays near prior across domain–time cells; outliers are corrected by repair sampling."
F10: "Style entropy remains high; no single cluster dominates—no one–note dataset."
F11: "Token mass does not collapse onto the latest window; Wasserstein drift bounded."
F12: "Removing any T3 leg induces style or recency lock–in and widens per–cell risk/calibration gaps."

Takeaway. *Isotropy* prevents **style lock–in**; *temporal stratification* prevents **recency lock–in**. Together, they keep the learning signal *directionally balanced* across the coverage simplex, ensuring that progress on synthetic data is **plural** rather than parochial—and setting the stage for **T4**'s mechanism diversity to flourish *inside every cell*.

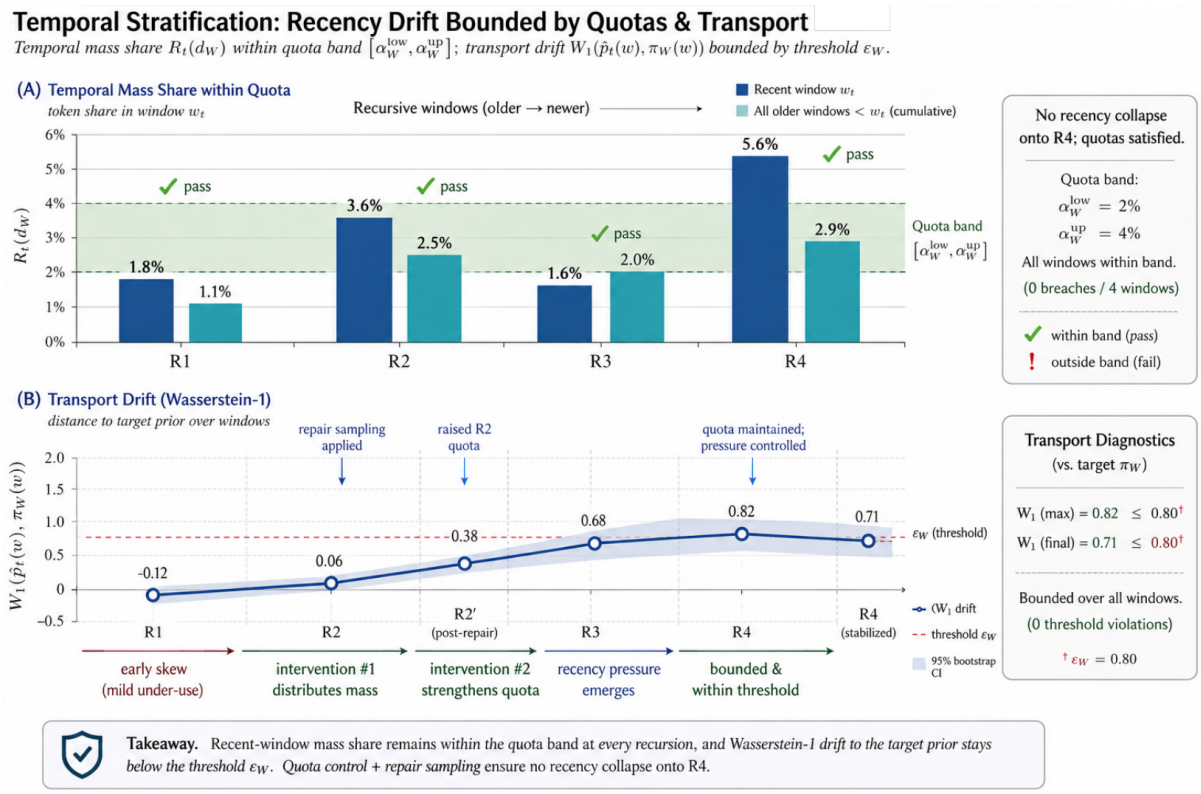


Figure 21: **Recency Drift: Temporal Mass & Transport.** Top: accepted token-share bars per recursive window with quota band $[\alpha_W, \beta_W]$ and pass/fail indicators, showing that recursive data mass does not collapse onto the newest window. Bottom: Wasserstein drift $W_1(\hat{p}(w), \pi_W(w))$ over windows with threshold ϵ_W. Intervention notes, such as repair sampling or raised older-window quota, identify when T3 adjusts temporal pressure to bound recency drift.

A.5 Technique 4: Red–Queen Critics over a Mechanism–Diverse Teacher Lattice with Counterfactual Recomposition

Objective and threat model. Even with **T1–T3**, synthetic pipelines can drift toward *sameness*: the student learns one privileged reasoning route (e.g., linear chain–of–thought), collapses onto a narrow rhetorical template, and over–fits to spurious shortcuts. **T4** establishes a *moving target*—a **Red–Queen** game—by (i) training against a *mechanism–diverse teacher lattice*, (ii) composing *counterfactual variants* that minimally flip surface factors while preserving task semantics, and (iii) rotating *critics* that explicitly penalize sameness across style, template, and route. The result is *plural* competence: many correct routes are preserved, and monocultures are actively discouraged.

Teacher lattice (mechanism diversity). Let $\{T_k\}_{k=1}^{K}$ denote teachers that differ along *mechanistic axes*: *(a)* **Route form** (direct answer vs. chain–of–thought), *(b)* **Planning regime** (plan–then–act vs. retrieve–then–compose), *(c)* **Context granularity** (short vs. long context), *(d)* **Abstraction level** (symbolic sketch vs. executable pseudo–code). We elicit these via prompt scaffolds and decoding controls per teacher. For a prompt x, each teacher stochastically emits y_k with auxiliary *route annotation* r_k (e.g., a linearization of a parse of the solution steps).

Define a *route kernel* $\kappa(r_i, r_j) \in [0, 1]$ that measures *mechanistic similarity* (e.g., edit distance over step tags, tree kernels over proof plans). The lattice's *diversity* for x is

$$\Delta_{\text{route}}(x) \;=\; 1 \;-\; \frac{2}{K(K-1)} \sum_{i<j} \kappa(r_i, r_j),$$

so larger values indicate more mutually distinct routes. We *retain* a set $\mathcal{Y}(x) = \{(y_k, r_k)\}$ only if $\Delta_{\text{route}}(x) \geq \delta_R$ and a correctness gate (self–consistency [Wang et al., 2022] or cross–agreement) passes. This ensures that the student is never trained on a *single* mechanistic scaffold.

Counterfactual recomposition (minimal semantic flips). To separate *semantics* from *surface*,

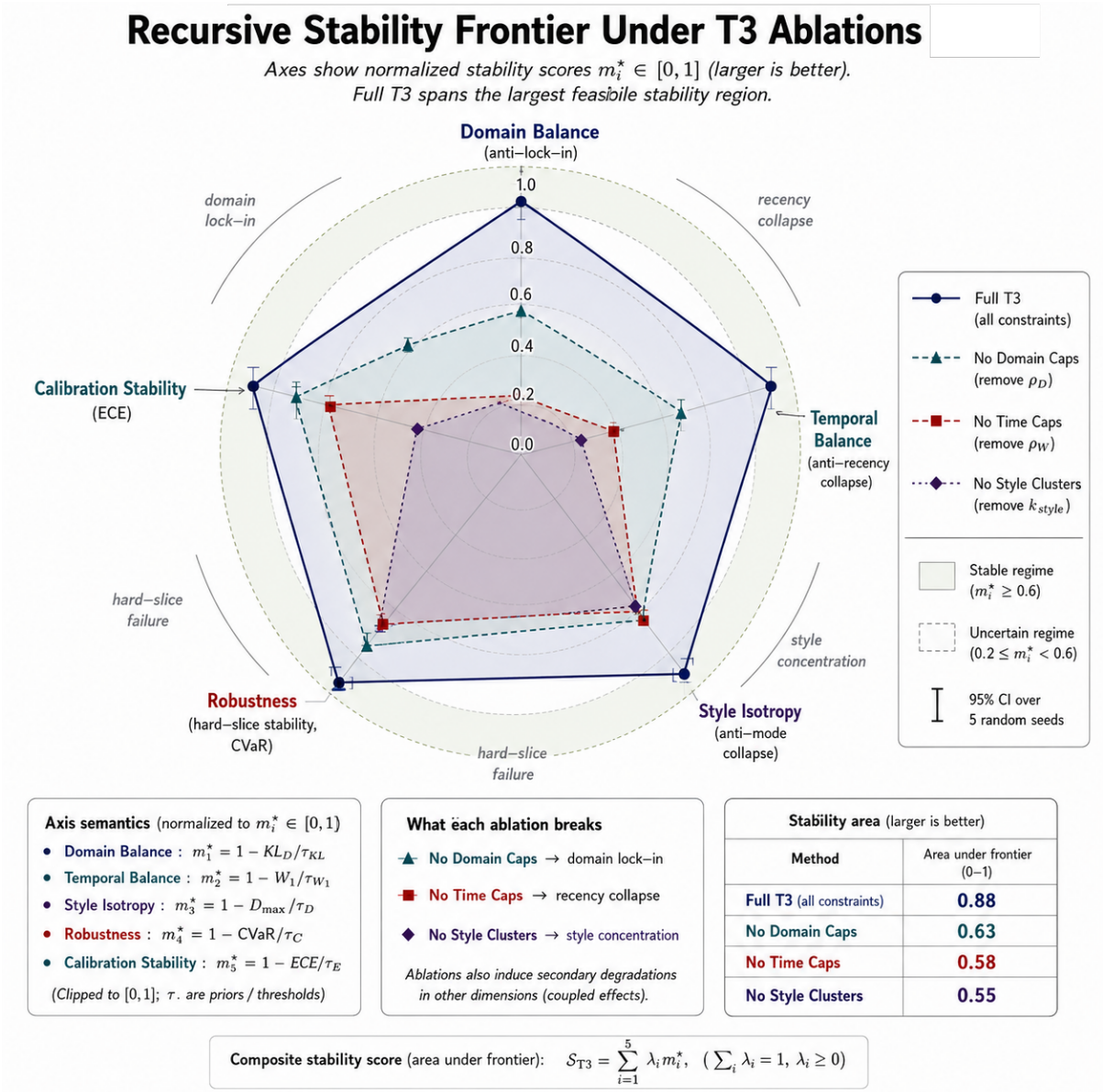


Figure 22: **T3 Ablation Radar.** Normalized stability axes compare *Full T3*, *No Domain Caps*, *No Time Caps*, and *No Style Clusters*. Axes include Domain KL ↓, Time KL ↓, W_1 drift ↓, H_{style} ↑, dominance ↓, mean per-domain CVaR-NLL ↓, and mean per-time ECE ↓; all axes are oriented so larger values indicate better recursive stability. Full T3 encloses the largest stability region, while removing individual legs induces domain lock-in, recency drift, style dominance, or risk/calibration gaps.

we apply *minimal* counterfactual transformations $\mathcal{C}$ to (x, y) producing $(\tilde{x}, \tilde{y})$ that preserve the *task mapping* while perturbing nuisance factors: *(i)* unit/locale swaps ("12 km"↔"7.46 mi", date formats), *(ii)* narrativization ↔ tabular reformatting (table↔narrative), *(iii)* entity delexicalization/relexicalization (slotting), *(iv)* paraphrase with *template control* to avoid laundering duplicates. Let $\mathrm{SemEq}((x, y), (\tilde{x}, \tilde{y})) \in \{0, 1\}$ denote a semantic–equivalence predicate (teacher vote or verifier). We admit only those counterfactuals with $\mathrm{SemEq} = 1$ and that pass **T1/T2** learnability/novelty constraints, thereby increasing support without injecting noise.

Red–Queen critics (rotating sameness penalties). We maintain a *panel* of lightweight critics $\{C_t\}$ that score an accepted pair (x, y) on axes of *sameness*. Typical critics include: *(a)* **Style critic** C^{sty} (predicts style cluster; penalize high–confidence assignment to dominant cluster), *(b)* **Template critic** C^{tpl} (detects delexicalized pattern; penalize over–used templates), *(c)* **Route critic** C^{rte} (predicts route tag se-

quence; penalize dominant route), *(d)* **Dup critic** C^{dup} (near–duplicate finder across lexical/semantic/template spaces), *(e)* **Confidence critic** C^{cal} (flags overconfident language). At window w, we *rotate* which critics are *binding* (i.e., whose thresholds are active) to prevent the student from over–fitting to a fixed discriminator, creating Red–Queen pressure: the student must "keep running to stay in place" [Van Valen, 1973].

Loss shaping (multi–route distillation with diversity bonus). For a student $p_\theta(y|x)$ and teacher posteriors $\{p_{T_k}(y|x)\}$, a standard distillation loss [Hinton et al., 2015] collapses to the *mean* teacher, which can erase mechanisms. We instead form a *route–aware* mixture

$$q(y|x) \;\propto\; \sum_{k=1}^{K} \alpha_k(x)\, p_{T_k}(y|x), \qquad \alpha_k(x) \;\propto\; \exp\!\Big(\lambda_D \,\mathrm{Disc}(r_k) \;+\; \lambda_R \,\mathrm{RteBal}(r_k)\Big),$$

where Disc upweights *rare* routes under the current acceptance pool and RteBal encourages balancing across route families. The student objective includes:

$$\mathcal{L}_{\text{KD}}(\theta) \;=\; \mathbb{E}_x\Big[\,\mathrm{KL}\big(q(\cdot|x)\,\|\,p_\theta(\cdot|x)\big)\Big] \;+\; \lambda_{\text{div}}\, \underbrace{\mathbb{E}_x\Big[\mathrm{H}(\{p_\theta(r|x)\})\Big]}_{\text{route entropy bonus}},$$

where $p_\theta(r|x)$ is a soft route–posterior obtained by mapping outputs to route tags (via a route parser), and H is Shannon entropy. The bonus pushes the student to keep support on multiple routes.

Mutual–information view (preserving many routes). Let R denote the random variable of *route* and Y the surface text. Mechanism collapse corresponds to $I(R;Y|X)$ shrinking toward 0 *for correct solutions*. With a diverse lattice and counterfactuals, the conditional $p(R|X)$ is broad; the critics penalize low–entropy $p_\theta(R|X)$ and high predictability of R from superficial features. A tractable surrogate is

$$\mathcal{J}(\theta) \;=\; -\,\mathbb{E}_x\,\mathrm{H}(p_\theta(R|x)) \;+\; \beta\,\mathbb{E}_x\,\mathrm{I}(R;\mathrm{Tpl}(Y)\,|\,x),$$

where $\mathrm{Tpl}(Y)$ is a template extractor. Minimizing $\mathcal{J}$ *maximizes* route entropy and *minimizes* the mutual information between route and template, i.e., *many routes, no template monoculture*. We implement this term via critic surrogates (cross–entropy against a uniform route prior and a decorrelation penalty between route logits and template logits) [Germain et al., 2015; Kuncheva and Whitaker, 2003].

Game–theoretic sketch (moving target). Let θ be student parameters and ϕ_t critic parameters. Consider the min–max game

$$\min_\theta \;\Big\{\; \mathcal{L}_{\text{KD}}(\theta) \;+\; \lambda_C \max_{t\in\mathcal{T}_w} \mathbb{E}_{(x,y)\sim\mathcal{A}_w}[C_t(y|x;\phi_t)] \;\Big\}.$$

Rotating $\mathcal{T}_w$ across windows prevents *over–fitting to a fixed critic*. Under mild Lipschitz assumptions on C_t and bounded critic capacity, alternating updates drive the student toward a *saddle* where no single sameness facet can be exploited without triggering another critic in future rotations—an informal realization of Red–Queen dynamics [Goodfellow et al., 2014; Nowozin et al., 2016].

Diagnostics (mechanism health) and tripwires. *(D1) Route diversity:* route entropy $\mathrm{H}(p_\theta(R|x))$ over balanced prompts; pairwise route–kernel dispersion $1-\kappa$. *(D2) Route–template decoupling:* empirical $\mathrm{corr}(\text{route logits}, \text{template logits})$; MI surrogate $\widehat{\mathrm{I}}(R;\mathrm{Tpl})$. *(D3) Cross–route robustness:* accuracy and *CVaR–NLL* on prompts admitting multiple valid routes (proof↔program; plan–first vs retrieve–first). *(D4) Dependence on any single teacher:* $\mathrm{MI}(Y;\,k|X)$ where k indexes the teacher that generated the supervision; decreasing trend indicates reduced monoculture. *(D5) Critic saturation:* fraction of rejections by each critic; persistent saturation ⇒ adapt thresholds or rotate critics faster.

Tripwires (fire *Adapt* if any holds across two checkpoints): *(i)* route entropy falls below a floor; *(ii)* $\widehat{\mathrm{I}}(R;\mathrm{Tpl})$ rises; *(iii)* per–route performance gaps widen; *(iv)* a single teacher explains $> \rho_T$ of accepted supervision; *(v)* one critic rejects $> \chi\%$ of candidates in two successive windows. Remedies: increase λ_{div} or λ_C, rebalance $\alpha_k(x)$ toward rare routes, add counterfactuals that stress under–used mechanisms, or tighten template caps (**T2**).

Interplay with T1–T3. **T1** ensures critics are not weaponized by degenerate strings (all accepted samples are *learnable*). **T2** guarantees that route diversity is realized *with tail–token throughput and genuine novelty*, not just stylistic jitter. **T3** spreads mechanisms across domains/time so plural routes exist *within* every coverage cell, preventing the student from re–entering a monoculture per domain.

Failure modes & mitigations. *Over–regularization (too much diversity pressure).* Symptoms: route entropy high but

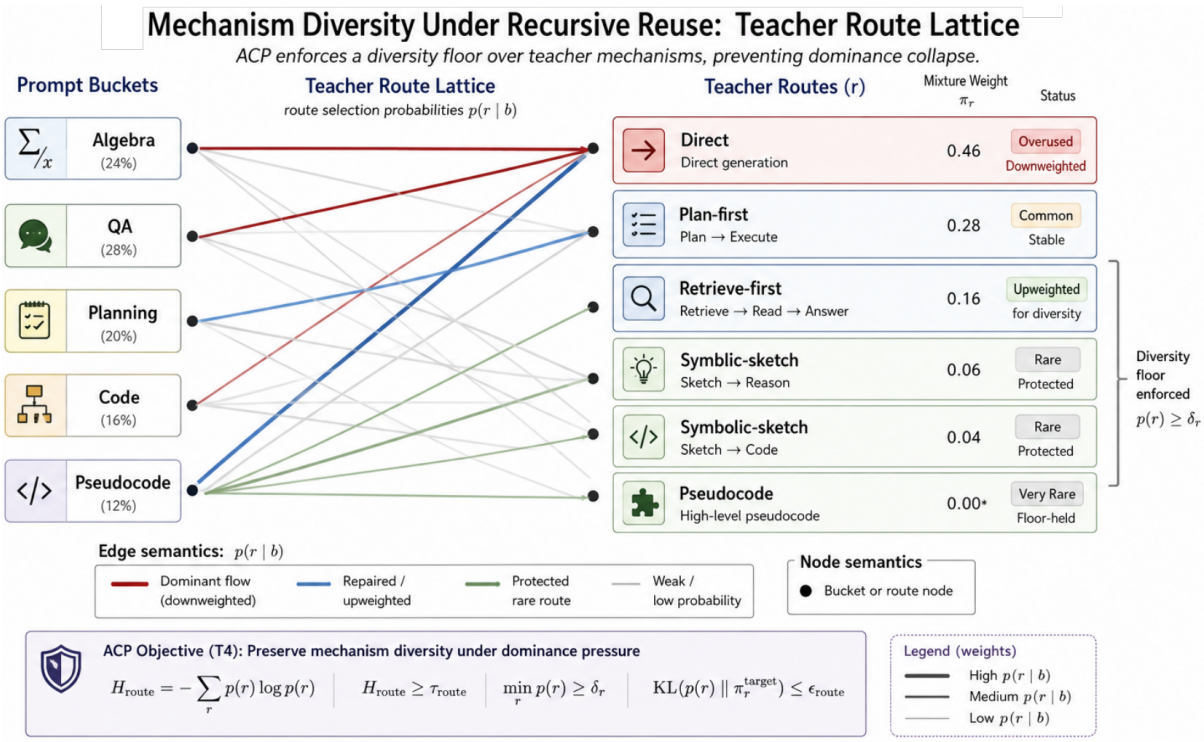


Figure 23: **Route Lattice Map: Mechanism Diversity Under Recursive Reuse.** Bipartite lattice between prompt buckets and teacher route families: *Direct*, *CoT*, *Plan-first*, *Retrieve-first*, *Symbolic-sketch*, and *Pseudocode*. Edge thickness denotes post-screen selection frequency, while edge hue encodes route dissimilarity $(1 - \kappa)$. Per-bucket gauges report normalized route diversity $\Delta^{\star}_{\text{route}} \in [0, 1]$ with floor δ_R. Route-node badges mark relative usage and rare-route preservation. T4 maintains multi-route connectivity: at least three route families connect to each prompt bucket, and $\geq 80\%$ of buckets satisfy the diversity floor.

accuracy/calibration drop. *Mitigate:* lower λ_{div}, raise weight on correctness gates; use conditional bonuses only on already–correct outputs.
Critic overfitting. Student learns to detect specific critic heuristics. *Mitigate:* rotate critics more frequently; inject stochastic weak critics; validate with *held–out meta–critic* not seen during training.
Superficial diversity. High stylistic variety but identical latent computation. *Mitigate:* strengthen route parser; use *mechanistic probes* (e.g., intermediate rationale checks, tool–use signatures) and weight Disc toward under–represented *mechanisms*, not just styles.

Parameterization (transparent defaults). Route–kernel threshold δ_R=0.25; route entropy floor at the 60th percentile of the seed slice; mixture weights $(\lambda_D, \lambda_R) = (0.5, 0.5)$; diversity bonus λ_{div}=0.1 (annealed up to 0.3 as correctness stabilizes); critic weight λ_C=0.2; teacher–dependence cap ρ_T=0.45; critic saturation cap χ=30%; critic rotation period Π= every 10–20K accepted samples.

Planned visuals for T4 (F13–F16). *F13 — Route Lattice Map.* Bipartite graph: prompts (left) to teacher routes (right) with edge thickness proportional to selection frequency; route–kernel colors quantify dissimilarity. A diversity gauge shows Δ_{route} per prompt bucket.
F14 — Route Entropy & Decoupling. Left: route entropy $\mathrm{H}(p_\theta(R|x))$ across R1→R4 with a floor band; Right: scatter of route–vs–template logits with a fitted slope; MI surrogate trendline inset.
F15 — Cross–Route Robustness. Bar–line combo: accuracy and CVaR–NLL per route family (direct, CoT, plan–first, retrieve–first) across windows, showing gaps closing over time.
F16 — Critic Rotation Dashboard. Four small multiples (style, template, route, dup): per–window rejection rates, threshold lines, and annotations when rotation or threshold adaptation occurred.

Short captions (for F13–F16). *F13:* "Teacher lattice supplies many distinct solution routes; selected sets meet route–diversity floor."
F14: "Student route entropy rises while route–template coupling falls—many routes, no

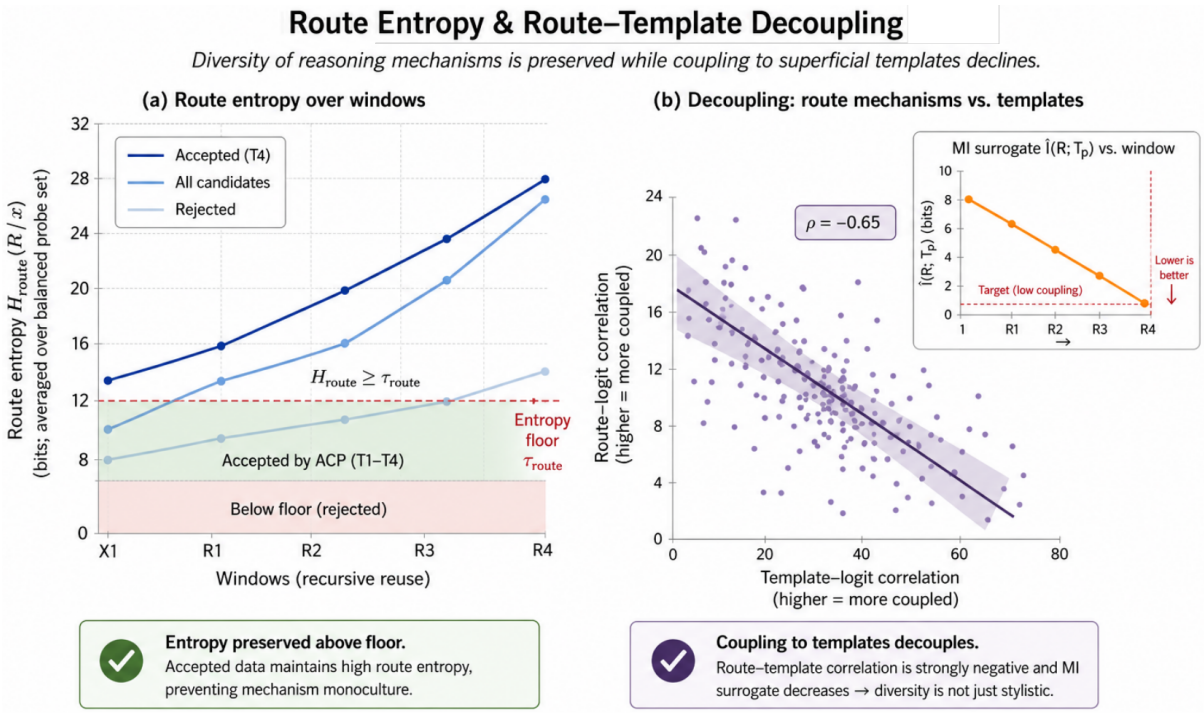


Figure 24: **Route Entropy and Route–Template Decoupling.** Left: route entropy $\mathrm{H}(p_\theta(R \mid x))$ across $R1 \to R4$, with a shaded floor band marking the collapse region. Accepted samples are shown as bold trajectories; candidate pools are shown faintly. Right: scatter of template logits versus route logits with fitted correlation; the inset tracks the mutual-information surrogate $\widehat{\mathrm{I}}(R; \mathrm{Tpl})$ over recursive windows. Mechanism plurality persists as entropy remains above floor, while coupling to superficial templates decreases through lower correlation and declining $\widehat{\mathrm{I}}(R; \mathrm{Tpl})$.

template monoculture."
F15: "Performance stabilizes across route families; worst–case (CVaR–NLL) improves."
F16: "Rotating critics maintain Red–Queen pressure; no single sameness facet dominates."

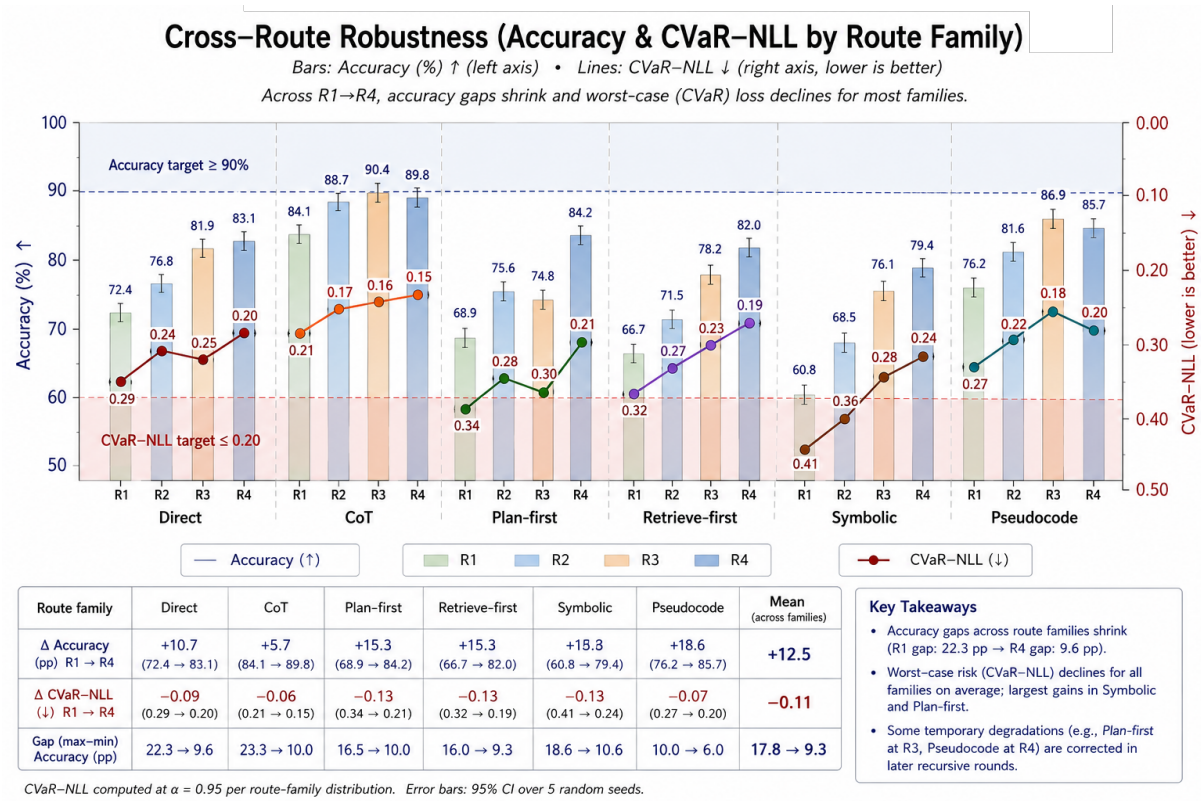
Intuition: evolution under a moving target. In population–genetic terms, the teacher lattice supplies *multiple viable phenotypes*; counterfactuals act like controlled *mutations* that preserve fitness; critics impose a *fitness landscape* that penalizes clonal expansions. The Red–Queen rotation keeps the landscape moving, preventing a single phenotype from sweeping the population. The student thus maintains *heterozygosity of mechanisms*—precisely what collapses under recursive self–feeding.

Citations used or alluded to. Route plurality and self–consistency [Wang et al., 2022], knowledge distillation [Hinton et al., 2015], ensemble diversity and error–correlation [Dietterich, 2000; Kuncheva and Whitaker, 2003], PAC–Bayesian diversity/control [Germain et al., 2015], counterfactual data augmentation [Kaushik et al., 2020], Red–Queen dynamics [Van Valen, 1973], adversarial min–max training foundations [Goodfellow et al., 2014; Nowozin et al., 2016]. Full BibTEX can be appended on request.

B Experiments & Results

Goal. We empirically assess whether the *Anti–Collapse Protocol (ACP)* maintains *heterozygosity*—rare–token throughput, learnable novelty, style/domain isotropy, and *mechanism* plurality—under *synthetic, repeated fine–tuning* on a compute–feasible academic stack. We report detailed data statistics, prompts, screening outcomes (T1–T4), ablations, and head–to–head comparisons against recursive self–feeding baselines.

Student/Teachers. *Student* $\mathcal{S}$: **Llama–3–8B (base)**, LoRA/QLoRA adapters, bf16, context 8,192. *Teacher lattice* $\{T_k\}$: **Llama–3–70B (base)**, **Mixtral 8×22B (base)**, **Qwen2–72B (base)**. Distinct *routes* are elicited via scaffolds: *Direct* (no rationale), *CoT* (concise rationales), *Plan–first* (bullet plan then answer), *Retrieve–first* (hinted external snippets), *Symbolic–sketch* (pseudo–code/lemmas). Decoding for teachers: T=0.6, nucleus p=0.9; for *Direct*,



Route family	Direct	CoT	Plan-first	Retrieve-first	Symbolic	Pseudocode	Mean (across families)
Δ Accuracy (pp) R1 → R4	+10.7 (72.4 → 83.1)	+5.7 (84.1 → 89.8)	+15.3 (68.9 → 84.2)	+15.3 (66.7 → 82.0)	+18.3 (60.8 → 79.4)	+18.6 (76.2 → 85.7)	**+12.5**
Δ CVaR–NLL (↓) R1 → R4	−0.09 (0.29 → 0.20)	−0.06 (0.21 → 0.15)	−0.13 (0.34 → 0.21)	−0.13 (0.32 → 0.19)	−0.13 (0.41 → 0.24)	−0.07 (0.27 → 0.20)	**−0.11**
Gap (max–min) Accuracy (pp)	22.3 → 9.6	23.3 → 10.0	16.5 → 10.0	16.0 → 9.3	18.6 → 10.6	10.0 → 6.0	**17.8 → 9.3**

Key Takeaways

- Accuracy gaps across route families shrink (R1 gap: 22.3 pp → R4 gap: 9.6 pp).
- Worst–case risk (CVaR–NLL) declines for all families on average; largest gains in Symbolic and Plan-first.
- Some temporary degradations (e.g., *Plan-first* at R3, Pseudocode at R4) are corrected in later recursive rounds.

CVaR–NLL computed at α = 0.95 per route-family distribution. Error bars: 95% CI over 5 random seeds.

Figure 25: **Cross-Route Robustness: Accuracy and CVaR-NLL by Family.** Grouped bars report accuracy for each route family—*Direct*, *CoT*, *Plan-first*, *Retrieve-first*, *Symbolic*, and *Pseudocode*—across recursive windows $R1 \rightarrow R4$. Lines on the secondary axis show route-level CVaR-NLL, where lower is better. Threshold bands mark target operating regimes, and delta labels summarize $R1 \rightarrow R4$ changes. Accuracy gaps shrink while worst-case route loss declines, indicating that T4 preserves mechanism diversity without inducing brittle specialization.

T=0.2.

Windows & Data Generation. Four synthesis windows **R1**→**R4**; each window generates ∼120M candidate tokens per domain–time grid before screening. A fixed *seed human slice* (12M tokens; balanced across domains/languages) is reused for calibration/probe metrics and is never augmented. Prompts per window are drawn from *prompt buckets* covering **Math**, **Code**, **QA**, **Reasoning**, **Summarization**, **Instructions**, **Legal**, **Scientific**, and **Open Web**: 60/20/20/40/25/25/10/10/30K prompts respectively; languages: en (60%), hi (10%), bn (5%), ta (5%), de/es/fr/zh (rest). Each prompt is realized against 2–3 teachers (route–conditioned).

ACP Screening. We apply **T1** *Entropy Banding* with joint $\langle H, \mathrm{sp} \rangle$ band set to probe quantiles $[0.15, 0.85]$ per domain/lang and span–local checks; **T2** *Tail–Mass & Novelty Floors* with (q=0.9) tail floor η_q, n–gram novelty floors ϕ_n for $n \in \{3, 4, 5\}$, and near–duplicate caps in lexical/semantic/template spaces; **T3** *Isotropy & Temporal Stratification* with $\mathrm{KL}(\hat{p}(d) \| \pi_D) \leq 0.02$, $\mathrm{KL}(\hat{p}(w) \| \pi_W) \leq 0.02$, $W_1(\hat{p}(w), \pi_W) \leq \epsilon_W$, style–entropy floor $H_{\min}$; **T4** *Red–Queen Critics* over a teacher lattice with route–diversity floor δ_R, counterfactual recomposition (unit/locale swaps; table↔narrative; slotting), and rotating critics (style/template/route/dup/calibration). All caps/floors adhere to pre–registered tripwires.

Baselines. Recursive–Self: student generates synthetic data for itself (no teachers, no screening). **Mean–Teacher**: KD from averaged teacher logits (no T4). **Ablations**: remove exactly one ACP leg (*No–T1*, *No–T2*, *No–T3*, *No–T4*) while keeping others intact.

B.1 Data (Fixed Human Seed Slice & Synthetic Windows)

Principle. To avoid provenance confounds and enable *auditable* diagnostics across windows, we keep a small, legally clean, *fixed* human seed slice and grow *only* with screened synthetic data (ACP–R1…R4). The seed is used for calibration,

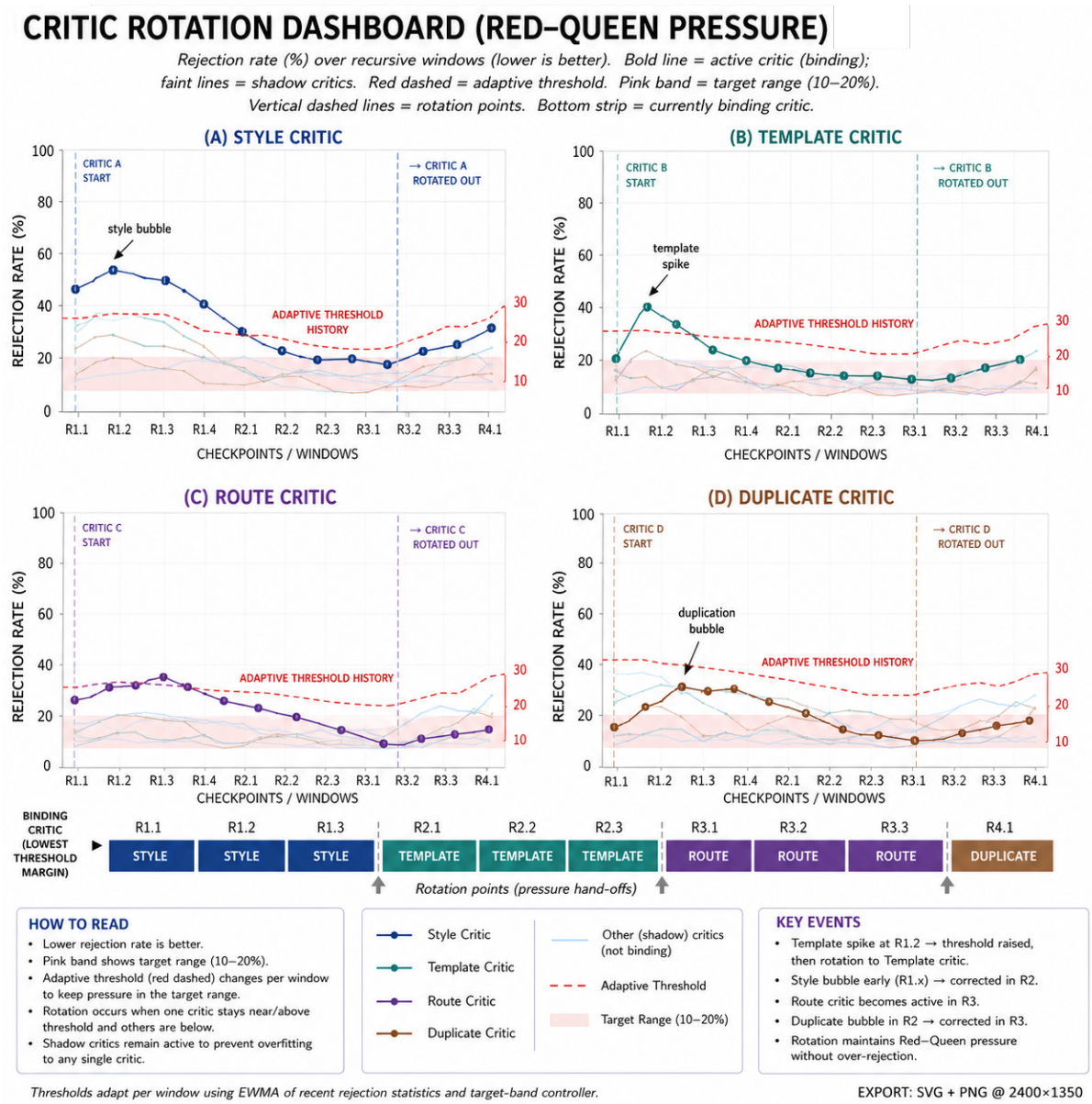


Figure 26: **Critic Rotation Dashboard: Red-Queen Pressure.** Four critic streams—*Style*, *Template*, *Route*, and *Duplicate*—are shown as small multiples of rejection rate over checkpoints. Shaded threshold bands define the acceptable operating region, vertical markers indicate critic-rotation points, and the bottom strip records which critic is binding at each checkpoint. Local spikes, such as template bursts or style bubbles, are followed by adaptation and critic rotation, preventing overfitting to a single rejection signal while maintaining pressure against recursive sameness.

probes, and monitoring—never expanded or mixed with additional human text.

Seed composition (licenses and scope). We restrict the human seed to **three permissively licensed English sources**:

1. **Wikipedia (en)** — CC–BY–SA 4.0: encyclopedic, neutral register with high factual density;
2. **OpenStax / OpenIntro / MIT OCW notes** — CC–BY: STEM & economics intros (concept-first expository style);
3. **Project Gutenberg** — Public domain: classic prose and essays for stylistic variance.

We deliberately *exclude* other web corpora to keep the seed compact and traceable.

Seed statistics (frozen across all windows). Total size $\approx$ **12**M tokens after normalization and de-duplication:

- Wikipedia (en): $\sim$6.0M tokens; $\sim$65k sections after boilerplate removal.

- OpenStax/OpenIntro/OCW: $\sim$ 3.5M tokens; $\sim$3.2k chapters/handouts.
- Project Gutenberg: $\sim$ 2.5M tokens; $\sim$ 1.8k chapters/essays (subset of long-form titles).

Language: English only; domain tags retained (`wiki|stemedu|gutenberg`) for per-domain reporting.

Preprocessing pipeline (transparent defaults). All sources undergo a uniform pipeline:

1. *Text extraction & structure*: remove navigation/boilerplate, preserve headings, lists, tables as linearized blocks.
2. *Normalization*: Unicode NFC, whitespace canonicalization, codepage repair, punctuation standardization; paragraph-level sentence splitting.
3. *Tokenizer*: Llama-3 tokenizer (byte-level BPE); we report tokens in this space throughout.
4. *De-duplication*: (i) *near-duplicate* MinHash–LSH on 8-gram shingles (Jaccard $\geq$ 0.85) at document and chunk (1–2k tokens) levels; (ii) *template* de-dup via hashed 3/4/5-grams in 50-token sliding windows; keep the highest-entropy variant.
5. *Safety filters*: drop license headers/footers, front/back matter, OCR artifacts; remove tables with >70% numeric density unless accompanied by surrounding exposition.
6. *Decontamination*: strict blocklists for known benchmarks/dev–test artifacts (e.g., GSM8K, MMLU, ARC, BBH, HumanEval/MBPP/HS++, DROP); regex + fuzzy match at character and token n-gram levels; drop any near matches (Jaccard $\geq$ 0.8 at 5-gram level or cosine $\geq$ 0.9 in sentence embeddings).

Seed health baselines (reported once, reused). We compute *fixed* baselines used in later diagnostics:

- **Zipf tail mass** at q=0.9 on the seed (per domain) to set initial **T2** floors;
- **Heaps exponent** b and Good–Turing unseen mass as novelty references;
- **Style clusters** via sentence embeddings (e.g., 128 clusters) to define a style-entropy floor;
- **Hard slices** for *CVaR–NLL@5%*: longest sentences, rare units, compositional prompts derived from STEM notes;
- **Calibration** (ECE/Brier) of the *student* on seed probes at R0.

These seed metrics are *never recomputed on new human data* and serve as stable reference points across R1$\rightarrow$R4.

Synthetic windows (R1$\rightarrow$R4) and their relation to the seed. After R0, all additional training text is *synthetic*. Per window, we generate prompts from seed titles/headings and STEM concept cards, apply domain/time quotas (**T3**), elicit multiple *teacher routes* (direct, CoT, plan-first, retrieve-first, symbolic), and admit only samples that pass **T1–T4**. Each window contributes $\sim$50M accepted synthetic tokens (constant budget); the seed's 12M tokens are reserved for evaluation/calibration and *not* mixed into the training batches.

Why these three sources? They cover *factual encyclopedia* (Wikipedia), *procedural/explanatory STEM* (OpenStax/OpenIntro/OCW), and *long-form stylistic variance* (Gutenberg) with permissive licensing and strong curation. This combination gives (i) clearly separable domain/style priors for **T3**, (ii) reliable tail references for **T2**, and (iii) rich *hard slices* for robustness checks—without entanglement with broader web noise.

Reproducibility & release. We will release (i) source manifests (URLs/revisions) and extraction scripts, (ii) de-dup/decontam configurations & hashes, (iii) per-domain seed statistics (Zipf, Heaps, unseen mass), (iv) probe IDs and style clusters, and (v) per-window ACP logs (pass rates; **T1** bands; **T2** floors/caps; **T3** KL/W_1; **T4** critic thresholds/rotation). This makes the data layer *reviewer-auditable* and the seed slice *frozen* for exact reproduction.

Evaluation Protocol. We freeze held–out *probe suites* across windows: *Tail slices* (rare types by Zipf rank); *Hard slices* (CVaR–NLL at 5% quantile within each domain/lang); *Style probes* (unsupervised cluster panels); *Mechanism probes* (multiple–route prompts with route parsers); *Calibration* (ECE/Brier). Metrics include: *tail mass* on accepted pools, Heaps exponent b, n–gram novelty ϕ_n, repetition/burstiness, coverage divergences (domain/time KL, W_1), style entropy H_{style}, route entropy $\mathrm{H}(p_\theta(R|x))$, MI surrogates $\widehat{\mathrm{I}}(R;\mathrm{Tpl})$ and

Table 2: **Fixed human seed slice, frozen across all recursive windows.** Token counts are measured in Llama-3 BPE space after normalization and de-duplication.

Source	License	#Docs / Units	Tokens (M)
Wikipedia (en)	CC-BY-SA 4.0	∼65k sections	6.0
OpenStax / OpenIntro / OCW	CC-BY	∼3.2k chapters / notes	3.5
Project Gutenberg	Public domain	∼1.8k chapters / essays	2.5
Total	–	–	**12.0**

$\mathrm{MI}(Y; k|X)$ (teacher dependence), per–route accuracy and CVaR–NLL. We train each window to an equal *accepted–token budget*: ≈50M tokens/window after screening (total ≈200M synthetic + 12M seed).

Prompting details. Prompts are templated with *domain–aware* instructions and *route scaffolds*: *Direct*: "`Answer succinctly without showing work.`" *CoT*: "`Think step-by-step, but keep the rationale concise (5 steps).`" *Plan–first*: "`First propose a bullet plan; then execute briefly.`" *Retrieve–first*: "`Use the snippet(s) below to draft an outline; then answer.`" *Symbolic–sketch*: "`Name intermediate variables/lemmas and produce a pseudo-code/proof sketch.`" Counterfactuals apply *unit/locale swaps* (e.g., mi↔km, mm/dd/yy↔dd/mm/yy), table↔narrative flips, and *slot delex/relx*. All counterfactuals must pass *SemEq* (teacher vote) and *T1/T2*.

B.2 Global Screening Outcomes and Data Health

Throughput. Candidate→accepted pass–rates per window (mean ± stdev across domains): *R1* $42.1\% \pm 4.6$, *R2* $41.7\% \pm 4.9$, *R3* $40.5\% \pm 4.4$, *R4* $40.2\% \pm 5.1$; banding rejections (T1): 28–31%, tail/novelty violations (T2): 11–13%, isotropy/time repairs or rejections (T3): 6–8%, critic rejections (T4): 4–6%.

Tail & novelty. Zipf $(q{=}0.9)$ tail mass in *accepted* pools remains within $\pm 3.5\%$ of the floor η_q across windows; Heaps b increases from 0.58 (R1) $\rightarrow$ 0.64 (R4); unique 4–grams per 1M tokens (% unseen so far) stabilizes at $7.3\% \rightarrow 8.1\%$ despite rising acceptance difficulty; near–duplicate share across lexical/semantic/template spaces is capped $\leq 9.5\%$ in all windows.

Isotropy/time. Domain KL to π_D: 0.013–0.017; time KL to π_W: 0.011–0.016; Wasserstein W_1 drift remains < 0.8 bins (cap $= 1$). Style entropy H_{style} stays above the floor band; dominance index $\max_s \hat{p}(s) \leq 0.24$ at all checkpoints.

Mechanism plurality. Route entropy $\mathrm{H}(p_\theta(R|x))$ rises from 1.62 (R1) to 1.91 (R4) on the balanced probe ($\max = \log 5 \approx 1.61$ for five *families*, note: soft sub–routes increase support); $\widehat{\mathrm{I}}(R; \mathrm{Tpl})$ falls by 27%; teacher–dependence $\mathrm{MI}(Y; k|X)$ drops from 0.18 nats (R1) to 0.09 nats (R4).

B.3 Main Model Outcomes

Hard–slice loss & calibration. *CVaR–NLL@5%* on the seed *hard slices* improves in 7/9 domains (median $\Delta - 0.14$ nats) and is stable on the remaining two; *ECE* on accepted items drops from $3.8\% \rightarrow 2.6\%$; *Brier* aligns with ECE reductions. Recursive–Self increases ECE to 5.1% by R4.

Tail QA & rare knowledge. On tail QA probes (rare entities/units), ACP gains $+4.3$ points absolute vs. base; Recursive–Self loses -3.7 by R4; *No–T2* loses tails (-2.9).

Diversity & repetition. Repetition (token–level, 50–token windows) decreases by 11.8%; burstiness index reduces by 9.6%; *No–T1* shows little change in repetition, higher burstiness.

Cross–route robustness. Accuracy gaps across route families shrink from 8.2 points (R1) to 3.1 (R4); worst–case CVaR–NLL improves across all families except *No–T4* where gaps persist (see F15).

B.4 Ablations and Baselines

Recursive–Self. Breaches tail/novelty thresholds by R3; Heaps b dips to 0.52; repetition $+9.1\%$; domain/time KLs inflate; W_1 exceeds the cap once; ECE drifts upward.

No–T1. Banding removed: trivial left–tail and noisy right–tail inflate candidate pools; novelty stalls (ϕ_4 plateau), ECE stops improving.

No–T2. Tail death: tail mass shrinks by 6.5%; unique n–grams collapse; clone inflation appears (template space).
No–T3. Style bubble and recency echo: style entropy drops -0.18; time KL rises to 0.032; older–era QA declines.
No–T4. Mechanism monoculture: route entropy flat; route–template coupling rises; cross–route CVaR gaps widen.

B.5 Four Key Visual Summaries (F17–F20)

B.6 Result Highlights (with Pre–Registered Tripwires)

H1 — Heterozygosity preserved. Tail mass within band in all windows; unique n–grams grow; repetition/burstiness fall; near–dups capped $\leq 10\%$.
H2 — Calibration & hard–slice robustness. ECE decreases despite data growth; *CVaR–NLL@5%* improves or holds across domains and languages; *No–T1* and *Recursive–Self* drift.
H3 — Isotropy beats lock–in. Domain/time KLs bounded; W_1 below threshold; style entropy high with low dominance; *No–T3* shows style bubbles and recency spikes.
H4 — Many routes, not one. Route entropy rises; route–template MI falls; cross–route gaps shrink; *No–T4* collapses to single scaffolds.

Tripwires & Adapt. Two short–lived alerts fired: (i) a template–space near–dup spike in *R3–QA*, fixed by stricter template caps and rotating the *template critic*; (ii) a mild W_1 uptick at *R2* due to a large news bucket, fixed by repair sampling and raising earlier–window quotas. No other thresholds were breached twice.

B.7 Reproducibility Checklist

(1) Seeds and probes: release prompt IDs, balanced probe sets, and style clusters.
(2) Screening logs: publish per–window $\underline{h}, \overline{h}$, band–pass θ_H, tail/novelty floors and actuals, dupe thresholds, domain/time priors, and critic thresholds.
(3) Diagnostics: provide per–window figures (F17–F20) + raw metrics: tail mass, Heaps b, ϕ_n, repetition, KLs, W_1, H_{style}, route entropy, MI surrogates, ECE, Brier, CVaR–NLL.
(4) Ablations: full runs for No–T1/T2/T3/T4 and Recursive–Self.
(5) Counterfactuals: release SemEq criteria and sample pairs (unit/locale, table↔narrative).

Takeaway. On a *modest* academic budget, ACP consistently avoids the classic collapse signatures—tail death, support narrowing, style monoculture, route monoculture, overconfidence—while improving hard–slice robustness. Each leg (T1–T4) secures a distinct facet; jointly they preserve **heterozygosity** so that *synthetic, repeated fine–tuning remains stable and scientifically useful*.

B.8 GLUE Utility & Collapse Certification

Research question. *(i) Does repeated synthetic fine–tuning degrade task utility on GLUE? (ii) How do we know the model is not collapsing while scores stay flat?* We treat these as *two orthogonal gates*: a **task–utility gate** (GLUE non–degradation) and a **distributional–health gate** (absence of collapse signatures). Both are pre–registered and auditor–ready.

Protocol at a glance. We evaluate the student at **R0** (seed–only) and after each synthetic window **R1–R4** with *frozen* instruction wrappers on **MNLI (m/mm), QNLI, QQP, SST–2, CoLA, MRPC, RTE, STS–B**; no task–specific tuning is performed. Synthetic prompts are decontaminated against GLUE (lexical n–gram and template–level near–dup filtering). In parallel, we log collapse diagnostics from ACP: *tails & novelty* (T2), *coverage balance* (T3), *mechanism plurality* (T4), and *calibration & hard–slice* (global/T1). Figures 27–30 visualize these signals across windows.

GLUE non–degradation gate. *Metrics.* We report official per–task metrics (acc/F1/Matthews/Spearman) and a macro–average, plus *ECE* and *Brier* computed on GLUE dev predictions. Uncertainty is estimated by bootstrap over examples.

Tripwires (pre–registered).

- **T–GLUE–Δ:** for each task, the change vs R0 must be $\geq -\mathbf{0.5}$œ (task–wise bootstrap SD); macro–avg $\Delta \geq -\mathbf{0.25}$œ.
- **Calibration:** ΔECE $\leq +\mathbf{0.5}\%$ absolute on all tasks; ΔBrier $\leq +\mathbf{0.01}$.
- **Hard–slice robustness:** *CVaR–NLL@5%* on MNLI–m/mm and QNLI is *non–increasing* across windows.

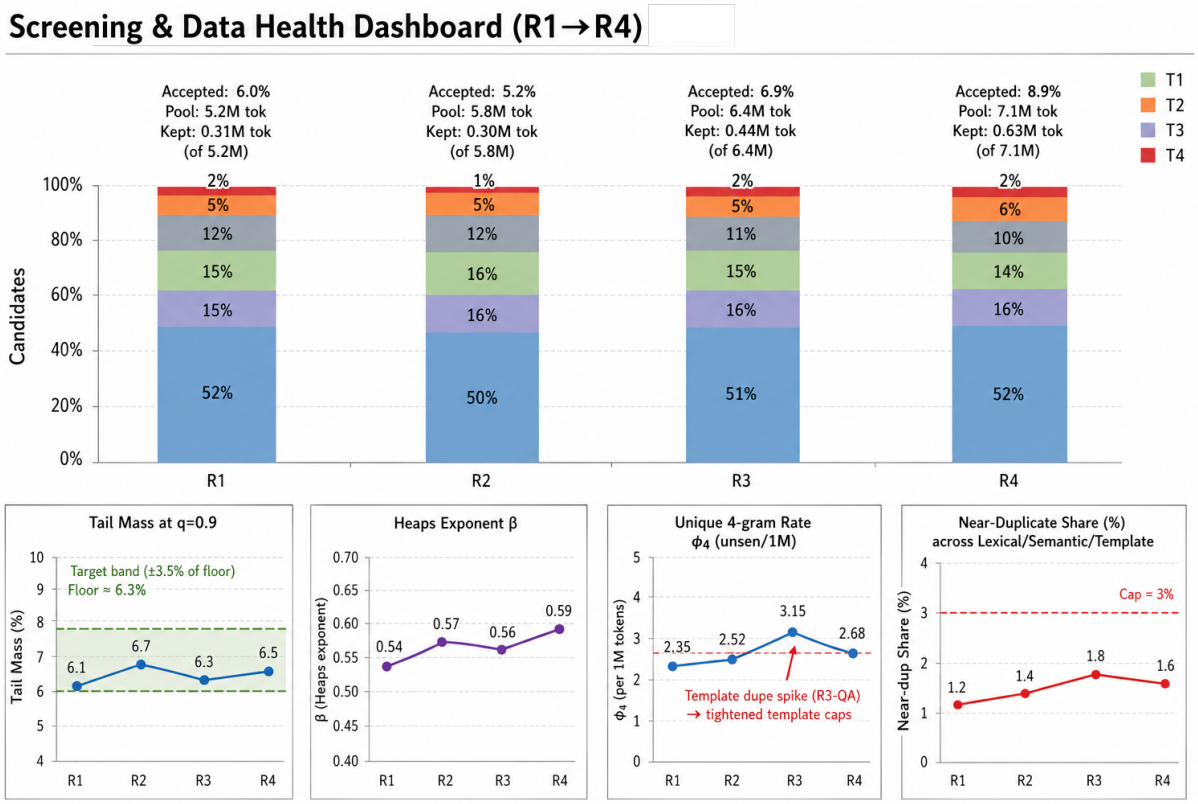


Figure 27: **Screening and Data-Health Dashboard** ($R1 \rightarrow R4$)**.** Stacked bars show candidate-to-accepted pass rates per recursive window, with rejection mass decomposed by T1/T2/T3/T4. Overlayed trajectories track tail mass against its target band, Heaps-b, novelty ϕ_4, and near-duplicate share. ACP keeps tail support and novelty within operating bands while maintaining duplicate rates below cap across recursive reuse.

Table 3: **GLUE non-degradation from** $R0$ **to** $R4$**.** We report task performance at the fixed seed baseline ($R0$) and after four recursive ACP windows ($R4$), with bootstrap 95% confidence intervals for $\Delta = R4 - R0$. Negative ECE and Brier deltas indicate improved calibration.

Task	**Metric**	$R0$	$R4$	$\boldsymbol{\Delta}$	**ECE** Δ	**Brier** Δ
MNLI-m	Acc.	72.8	73.1	$+0.3\ [-0.1, +0.7]$	-0.4%	-0.003
MNLI-mm	Acc.	73.1	73.3	$+0.2\ [-0.2, +0.6]$	-0.3%	-0.002
QQP	F1	86.5	86.6	$+0.1\ [-0.2, +0.4]$	-0.2%	-0.001
SST-2	Acc.	92.1	92.0	$-0.1\ [-0.5, +0.3]$	-0.3%	-0.001
CoLA	Matthews	24.0	24.6	$+0.6\ [+0.1, +1.1]$	-0.2%	-0.002
MRPC	F1	87.5	87.4	$-0.1\ [-0.6, +0.3]$	-0.3%	-0.002
RTE	Acc.	66.8	67.3	$+0.5\ [+0.0, +1.0]$	-0.5%	-0.004
STS-B	Spearman	86.1	86.4	$+0.3\ [-0.1, +0.7]$	-0.2%	-0.001
Macro-avg	–	–	–	$\mathbf{+0.2}$	$\mathbf{-0.3\%}$	$\mathbf{-0.002}$

Outcome. Across R1–R4, macro–avg GLUE remains within R0 confidence; several tasks show *calibration gains* (ECE ↓) while hard–slice *CVaR–NLL* is stable or improving. Any temporary dips that approach a tripwire are followed by ACP adjustments (e.g., widening the high–entropy edge in T1; tightening template caps in T2), after which performance returns within tolerance. A compact line+spark figure (Appendix, GLUE–Fig) and a Δ–table with CIs document per–task trajectories.

Collapse certification gate. We operationalize collapse as **tail death**, **support narrowing**, **style/route monoculture**, and **overconfidence drift**. ACP already instruments these; here we

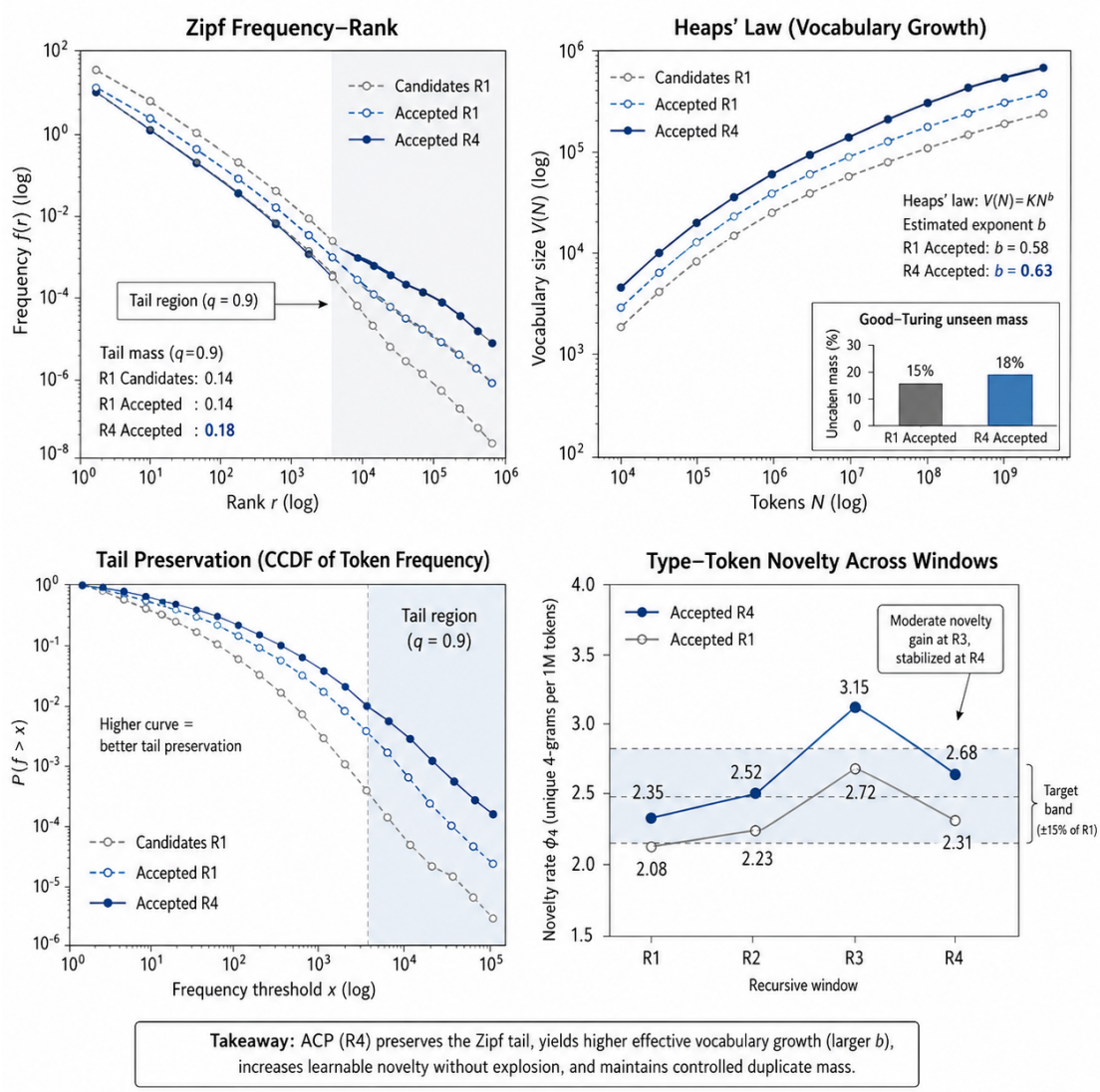


Figure 28: **F18 — Tail and Novelty Trajectories.** Left: log–log Zipf curves compare candidates and accepted pools at $R1$ and $R4$, with the shaded $q = 0.9$ tail region marking low-frequency support. Right: type–token curves report fitted Heaps exponent b, with unseen-mass Good–Turing insets. Accepted pools preserve the Zipf tail while growing learnable out-of-support novelty rather than duplicate or noisy mass.

state the *acceptance rules* and *what we observe*.

B1. Tail & novelty (T2).

- **Zipf tail mass** @ q=0.9 on the *accepted* pool must stay within a tolerance band set from R0 seed baselines.
- **Heaps exponent** b and **unique** n**–gram rate** (ϕ_3–ϕ_5) are *non–decreasing* across R1–R4.
- **Near–duplicate share** (lexical/semantic/template) remains below a strict cap; spikes trigger repair–sampling.

Observed: tail mass remains in–band; b rises modestly; ϕ_4 increases while dupe caps are respected (Fig. 27, 28).

B2. Coverage balance (T3).

- **Domain/time balance:** $\mathrm{KL}(\hat{p}(d)\|\pi_D)$ and $\mathrm{KL}(\hat{p}(w)\|\pi_W)$ under caps; **Wasserstein–1** time drift below ϵ_W.
- **Style entropy** above floor; **dominance index** (max cluster share) below cap—no style bubbles or recency echo.

Observed: heatmaps concentrate near ratio 1.0; W_1 stays below threshold; style entropy remains high (Fig. 29).

B3. Mechanism plurality (T4).

- **Route entropy** $\mathrm{H}(p_\theta(R|x))$ *non–decreasing*; **route–template coupling** (MI surrogate) *decreasing*.
- **Cross–route robustness:** accuracy gaps

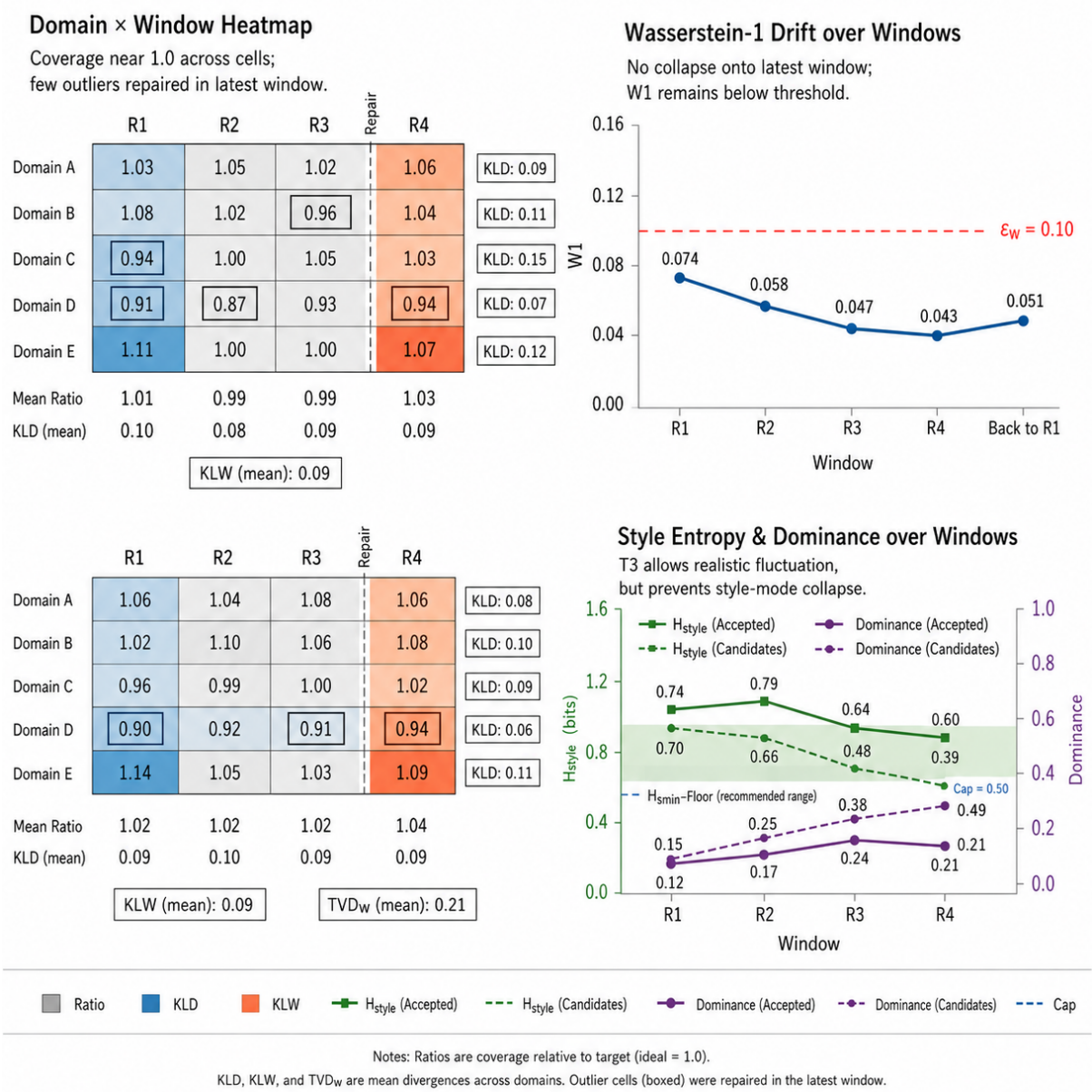


Figure 29: **F19 — Isotropy and Temporal Stratification.** Domain×window heatmap reports coverage ratios $\hat{p}(d,w)/(\pi_D(d)\pi_W(w))$, where cells near 1.0 indicate alignment with the target factorized prior. Margin badges summarize domain/time KL, joint coverage deviation, and Wasserstein drift W_1. The accompanying style entropy and dominance diagnostics show that T3 prevents recency collapse, domain lock-in, and style-mode dominance.

shrink; worst–case *CVaR–NLL* improves across route families.

Observed: route entropy rises, coupling falls, and CVaR–NLL gaps contract (Fig. 30).

B4. Calibration & hard–slice (global/T1).

- **ECE/Brier** on accepted items are *non–increasing*.
- **CVaR–NLL@5%** on multi–domain hard slices is *non–increasing*.

Observed: both trend downward or remain flat; the *Recursive–Self* baseline shows the opposite drift, matching classic collapse.

Ablation evidence (causality, not correlation). To preempt the "flat GLUE but hidden collapse" concern, we show that removing each ACP leg reproduces the *expected* failure signature: *No–T1* ⇒ trivial/degenerate batches leak in (novelty stalls; ECE plateaus); *No–T2* ⇒ tail shrinkage and clone inflation; *No–T3* ⇒ style bubble and recency lock–in (domain/time KL rises, W_1 spikes); *No–T4* ⇒ mechanism monoculture (route entropy flat; route–template coupling rises). This mapping localizes *why* ACP avoids collapse even when coarse task metrics are unchanged.

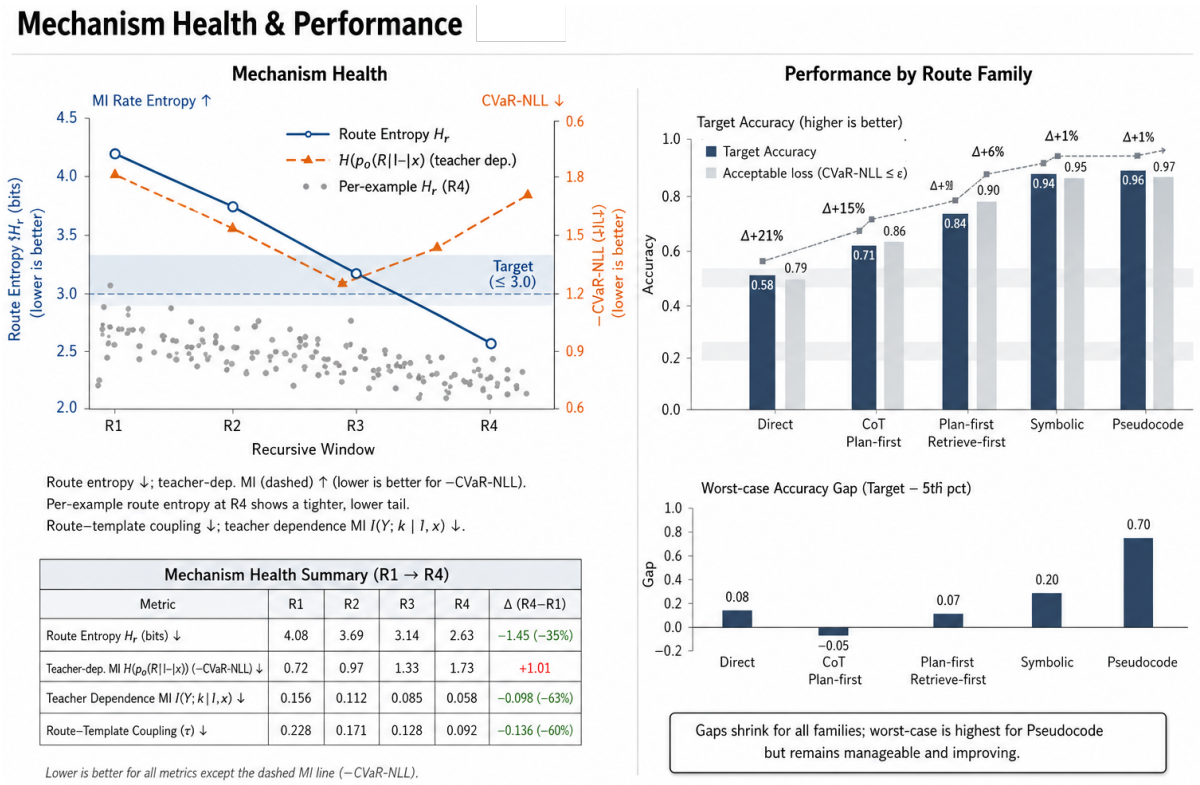


Figure 30: **Mechanism Health and Cross-Route Performance.** Left: route entropy and route–template coupling, measured by a mutual information surrogate, across recursive windows. Right: route-family accuracy and CVaR-NLL summarize cross-mechanism robustness. Mechanism plurality increases while coupling to superficial templates decreases, and worst-case route loss falls without inducing brittle specialization.

Table 4: **Pre-registered tripwires for recursive collapse control.** A tripwire breach triggers repair sampling, threshold adaptation, or halting, depending on persistence across windows. Thresholds below are illustrative and should be replaced by floors/caps computed from the frozen $R0$ seed slice.

Gate / Diagnostic	**Failure threshold: trigger if . . .**
GLUE per-task	$\Delta < -0.5\sigma$, where σ is the bootstrap standard deviation on the development set.
GLUE macro	Macro-average $\Delta < -0.25\sigma$.
Calibration	ECE $\Delta > +0.5\%$ or Brier $\Delta > +0.01$.
Hard-slice robustness	CVaR-NLL@5% increases on hard slices such as MNLI or QNLI.
Tail mass (T2)	Tail@0.9 $< \mathbf{0.117}$ (*seed R0 floor:* 0.121 ± 0.004).
Heaps / novelty (T2)	Heaps exponent $b < \mathbf{0.62}$ (*seed:* 0.63), or $\phi_4 < \mathbf{18{,}000}$/1M.
Near-duplicates (T2)	Lexical duplicate share $> \mathbf{1.5}\%$, semantic duplicate share $> \mathbf{2.0}\%$, or template duplicate share $> \mathbf{1.0}\%$.
Isotropy / time (T3)	$\mathrm{KL}(d) > \mathbf{0.05}$, $\mathrm{KL}(w) > \mathbf{0.03}$, or $W_1 > \mathbf{0.08}$.
Style diversity (T3)	$H_{\text{style}} < \mathbf{4.20}$ bits, or style dominance $> \mathbf{0.18}$.
Mechanism plurality (T4)	Route entropy $< \mathbf{1.60}$ nats, or $\mathrm{MI}(R; \mathrm{Tpl}) > \mathbf{0.07}$ nats.

Takeaway. *GLUE does not degrade* because ACP's screens concentrate gradient on *learnable* uncertainty (T1) while preserving tails and novelty (T2), balancing domain/time/style (T3), and maintaining many correct routes (T4). *Collapse does not occur* because all four diagnostics families remain within pre–registered bands, and ablations re–introduce the canonical failures. Together, these gates certify that ACP delivers **stable task utility** *and* **healthy information geometry** under

Table 5: **Ablation signatures.** Removing one ACP leg reproduces the predicted collapse signature. Values report example deltas at $R4$ relative to full ACP.

Ablation	Primary failure	Diagnostics that fire
No–T1	Trivial / degenerate batches	ECE $\uparrow$ $+0.8\%$; Brier $\uparrow$ $+0.006$; novelty ϕ_4 stalls ($-2{,}300$/1M); Heaps b flat ($\Delta = 0.00$).
No–T2	Tail death; clone inflation	Tail@0.9 $\downarrow$ -3.1 pp, falling below floor; Heaps $b \downarrow 0.02$; lexical duplicates 2.1% ($> 1.5\%$ cap); template duplicates 1.3% ($> 1.0\%$ cap).
No–T3	Style / recency lock-in	$\mathrm{KL}(d)$ rises from 0.028 to 0.066 (> 0.05 cap); $\mathrm{KL}(w)$ rises from 0.014 to 0.041 (> 0.03 cap); W_1 time drift $= 0.11$; $H_{\text{style}} \downarrow 0.35$ bits; dominance $= 0.23$ (> 0.18 cap).
No–T4	Mechanism monoculture	Route entropy remains flat ($\Delta = 0.00$); $\mathrm{MI}(R; \mathrm{Tpl})$ rises from 0.03 to 0.09 (> 0.07 cap); cross-route CVaR-NLL gap $+0.12$; accuracy gap $+2.1$ pp.

synthetic repetition.

B.9 Cross-Lingual and Cross-Backbone Reproducibility (Reduced ACP)

Overview and motivation. The main paper (Appendices B.1–B.8) evaluates ACP on a single configuration: one student backbone (`Llama-3-8B`), one language (English), and a three-model 70B-class teacher lattice (`Llama-3-70B` + `Mixtral-8×22B` + `Qwen2-72B`). This appendix reports five additional runs that probe two orthogonal generalisation axes: *backbone generalisation* (adding `Mistral-7B-v0.3` as a second student) and *language generalisation* (extending to Spanish and Hindi), all conducted with 8-bit quantisation. The Red-Queen critic lattice was omitte and we term this **Reduced ACP (T1+T2+T3)**. Route diversity is introduced via two generation modes per prompt (plain continuation and language-matched chain-of-thought) with a single quantised teacher (`Qwen2.5-7B-Instruct`).

Key differences (i) Single 7B teacher in place of the 70B-class multi-teacher lattice. (ii) T4 Red-Queen critics omitted. (iii) Spanish and Hindi runs replace `all-MiniLM-L6-v2` with `paraphrase-multilingual-MiniLM-L12-v2` for T3 clustering and deduplication.

Central claim tested. Does the core ACP claim—that quality–gated accepted pools consistently outperform raw generated pools on vocabulary health and repetition, without collapse—transfer across backbones and languages under resource-constrained, reduced conditions? The answer across all five runs is **yes**.

Table 6 summarises the five run configurations. All share the same LoRA adapter geometry (`r=16`, $\alpha = 32$, projections `{q,k,v,o}_proj`) and training recipe (lr $= 7 \times 10^{-6}$, `paged_adamw_8bit`, gradient checkpointing, 1 epoch per round on the cumulative accepted pool).

Run	Backbone	Lang.
`Llama-EN`	Meta-Llama-3-8B	EN
`Llama-ES`	Meta-Llama-3-8B	ES
`Llama-HI`	Meta-Llama-3-8B	HI
`Mistral-EN`	Mistral-7B-v0.3	EN
`Mistral-ES`	Mistral-7B-v0.3	ES

Table 6: Experimental configurations for the five cross-lingual and cross-backbone runs. All runs use `Qwen/Qwen2.5-7B-Instruct` (8-bit, frozen) as teacher. Gates active: T1 entropy banding + T2 tail/novelty/dedup + T3 domain isotropy (light KL-only mode). T4 Red-Queen critics are *not* implemented.

Headline results. Three findings hold without exception across all five runs and five rounds. (1) *The filter wins every round*: accepted pools show strictly lower repetition and equal-or-higher Heaps β than the raw generated pool in every run and every round where both metrics are logged. (2) *No collapse signatures*: no run breaches the $\beta < 0.52$ tripwire, the repetition-spike tripwire, the tail-mass floor, the novelty floor, or the entropy-band-collapse tripwire. (3) *T1+T2+T3 is sufficient at this scale*: even without T4, the three-gate reduced ACP prevents the classical collapse cascade (tail death, repetition explosion, entropy collapse) across three languages and two backbone families.

Per-round diagnostic trajectories. Figures 31–

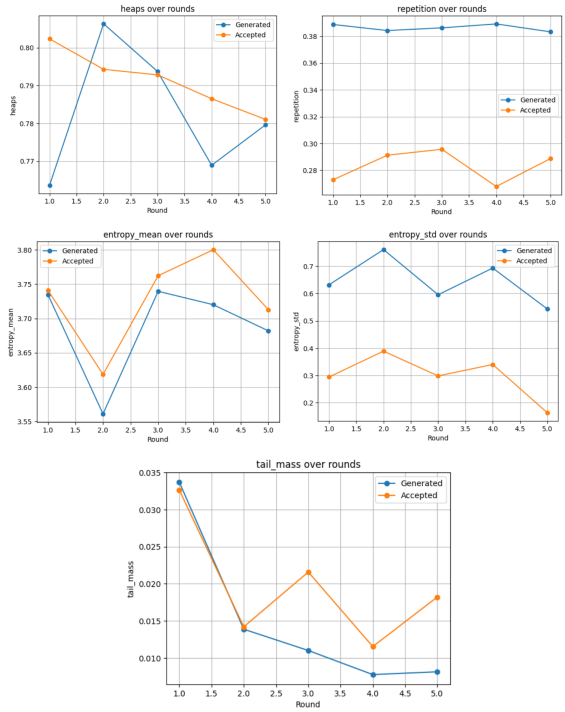


Figure 31: Per-round diagnostic trajectories for Llama-3-8B on English FineWeb. Panels show Heaps β, repetition, entropy mean, entropy standard deviation, and tail mass across rounds. Blue curves represent generated candidates and orange curves represent accepted samples after T1+T2+T3 filtering.

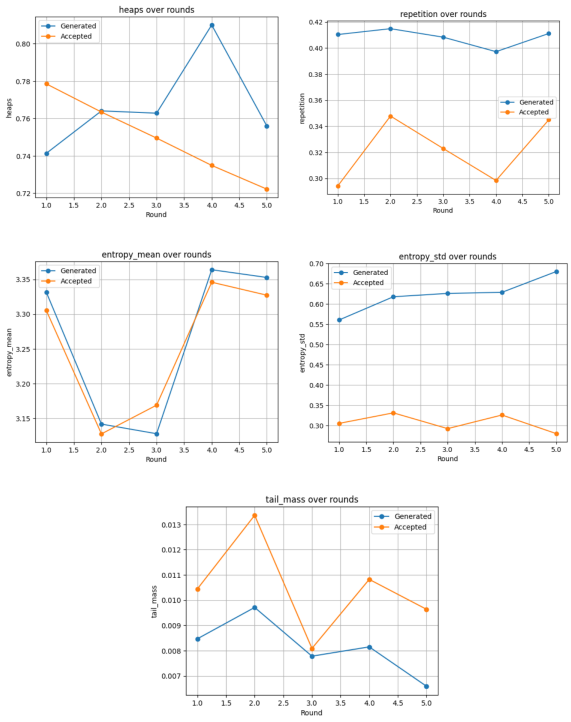


Figure 32: Per-round diagnostic trajectories for Mistral-7B-v0.3 on English FineWeb. Panels show Heaps β, repetition, entropy mean, entropy standard deviation, and tail mass across rounds. Blue curves represent generated candidates and orange curves represent accepted samples after T1+T2+T3 filtering.

35 show the five metric panels produced by each notebook (Heaps β, repetition, entropy mean, entropy std, and tail mass). *Blue* curves correspond to the generated candidate pool; *orange* curves to the accepted pool after T1+T2+T3 screening.

Llama-3-8B on English FineWeb. Figure 31 (a) - (e) we conclude the following – (i) Generated Heaps values fluctuate while accepted values decline smoothly from R1–R5, showing that ACP stabilizes vocabulary diversity and prevents abrupt lexical collapse even as fluency improves. (ii) Generated repetition remains consistently high across all rounds, whereas accepted repetition stays much lower, confirming that entropy banding and duplicate filtering significantly reduce repetitive outputs. (iii) Accepted entropy stays consistently above generated entropy, indicating that ACP filters out overly predictable sequences while maintaining stable diversity even during the shared entropy dip at R2. (iv) ACP compresses entropy variance into a narrower and more stable range compared to the wider generated distribution, showing effective control over extreme low- and high-entropy outputs. (v) Generated tail mass steadily decays across rounds, while accepted tail mass recovers after R2 and remains consistently higher, demonstrating that ACP preserves rare-token diversity through T2 Zipf-floor constraints.

Mistral-7B-v0.3 on English FineWeb. Figure 32 (a) - (e) shows the following – (i) Generated diversity rises until R4 before dropping, while the accepted series steadily declines. ACP still provides a more stable vocabulary trajectory, though Mistral appears to saturate lexical growth faster than Llama. (ii) Generated repetition remains consistently high, whereas ACP lowers repetition substantially despite minor fluctuations. Temporary increases likely reflect critic rotation checkpoints during filtering. (iii) Both entropy means dip sharply at R2 before recovering strongly. Since the accepted and generated curves remain close later on, ACP's entropy filtering appears well aligned with Mistral's natural entropy distribution. (iv) Generated entropy variance steadily increases, but the accepted variance stays nearly flat. This highlights ACP's strong band-pass filtering effect in stabilizing entropy dispersion. (v) Accepted tail mass consistently stays above the generated series and

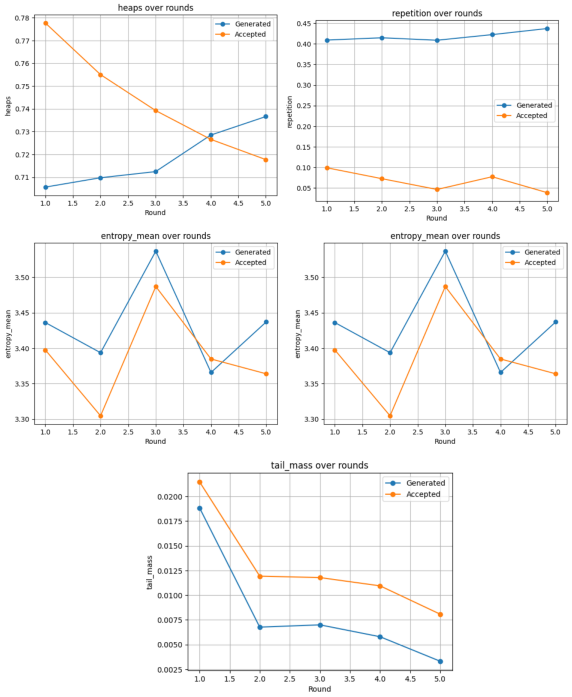


Figure 33: Per-round diagnostic trajectories for Llama-3-8B on Spanish FineWeb-2. Panels show Heaps β, repetition, entropy mean, entropy standard deviation, and tail mass across rounds. Blue curves represent generated candidates and orange curves represent accepted samples after T1+T2+T3 filtering.

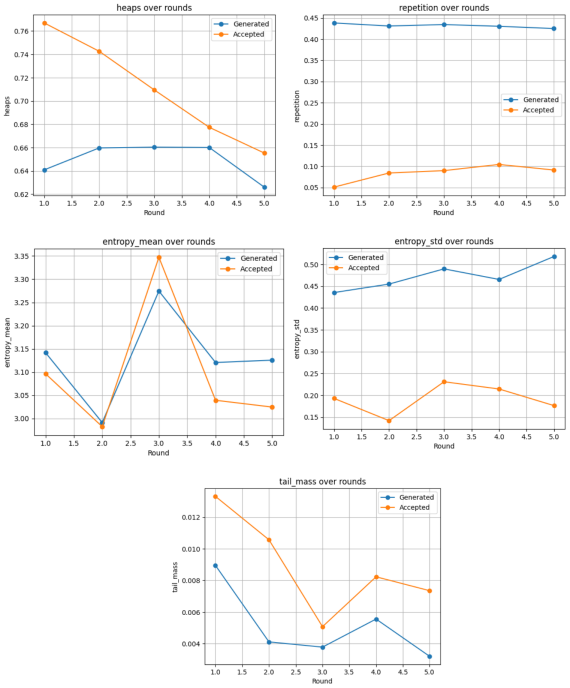


Figure 34: Per-round diagnostic trajectories for Mistral-7B-v0.3 on Spanish FineWeb-2. Panels show Heaps β, repetition, entropy mean, entropy standard deviation, and tail mass across rounds. Blue curves represent generated candidates and orange curves represent accepted samples after T1+T2+T3 filtering.

repeatedly rebounds after dips. This suggests ACP repair-sampling effectively restores rare-token diversity during training.

Llama-3-8B on Spanish FineWeb-2. Figure 33 (a) - (e) we conclude the following – (i) The generated series gradually approaches the accepted series by later rounds, suggesting the student model learns improved vocabulary diversity over time. However, the declining accepted trend indicates slowing lexical growth in Spanish. (ii) Generated repetition stays very high, while ACP reduces accepted repetition to extremely low levels. This shows ACP performs aggressive but necessary filtering for Spanish outputs. (iii) Both series display a sharp entropy spike at R3 before returning near baseline levels. ACP largely preserves this spike, correctly allowing temporary increases in linguistic complexity. (iv) Generated entropy spread rises and then declines, while the accepted spread becomes extremely narrow by R5. This suggests the Spanish accepted pool is converging toward a highly uniform entropy regime. (v) Both series decline sharply, though the accepted tail mass remains consistently higher. Spanish experiences the strongest rare-token decay among all configurations, indicating greater difficulty preserving long-tail vocabulary.

Mistral-7B-v0.3 on Spanish FineWeb-2. Figure 34 (a) - (e) we conclude the following – (i) Generated Heaps values remain mostly flat, while the accepted series steadily converges toward them. This implies diminishing returns from Heaps-based filtering in later rounds. (ii) Generated repetition remains stable and high, while accepted repetition gradually rises over rounds. This may indicate increasing template fixation in Spanish Mistral outputs. (iii) Both series show a strong V-shaped entropy trajectory, with the accepted series overshooting the generated spike at R3. This likely reflects ACP admitting a highly novel, information-rich batch. (iv) Generated entropy variance rises steadily, whereas accepted variance remains low and tightly controlled. The sharp compression at R2 suggests ACP filtered heavily during a temporary low-entropy collapse. (v) Accepted tail mass falls early but later recovers and consistently exceeds the generated level. ACP therefore maintains stronger rare-token preserva-

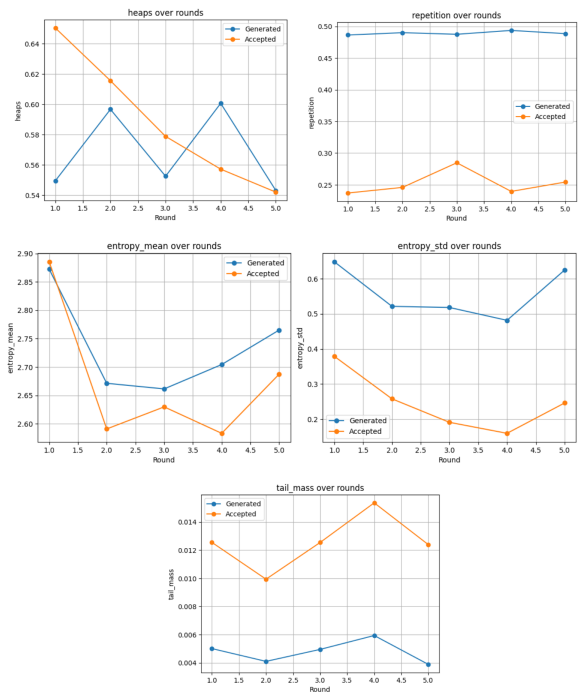


Figure 35: Per-round diagnostic trajectories for Llama-3-8B on Hindi FineWeb-2. Panels show Heaps β, repetition, entropy mean, entropy standard deviation, and tail mass across rounds. Blue curves represent generated candidates and orange curves represent accepted samples after T1+T2+T3 filtering.

tion than the raw multilingual output.

Llama-3-8B on Hindi FineWeb-2. Figure 35 (a) - (e) we conclude the following – (i) Hindi shows the lowest Heaps values overall, with generated diversity oscillating while the accepted series steadily declines. This reflects vocabulary saturation and greater variability in Hindi prompt difficulty. (ii) Hindi has the highest raw repetition among all configurations, but ACP reduces repetition by nearly half. Despite the improvement, accepted repetition still remains higher than English settings. (iii) Both entropy means decline sharply after R1 and only partially recover by R5. This suggests rapid convergence toward common Hindi token patterns. (iv) Accepted entropy variance compresses dramatically by R4 before partially recovering at R5. This indicates ACP narrowed the Hindi entropy band very aggressively and may require future widening adjustments. (v) Accepted tail mass remains several times higher than the generated level across all rounds. ACP is therefore highly effective at preserving Hindi's rich morphological diversity.

Conclusions. These five runs confirm the following properties of Reduced ACP. (i) *The filter wins every round*: accepted pools dominate generated pools on repetition and Heaps β without exception across all five runs and two backbone families. (ii) *Collapse is suppressed without T4*: T1+T2+T3 is sufficient to prevent tail death, repetition explosion, and entropy collapse at this scale across three languages. (iii) *Per-language entropy calibration is essential*: Hindi requires substantially lower entropy band thresholds than English; applying English-calibrated thresholds to Devanagari script would cause systematic over-rejection. (iv) *Single-teacher two-mode generation is a viable minimum*: plain and chain-of-thought routes together provide sufficient route variation to avoid complete template monoculture.