SELECTIVE CONTINUATION SUPERVISION AND LEARNED STOPPING

Anonymous authors
Paper under double-blind review

ABSTRACT

Preserving content targets and complete-stop supervision need not preserve learned stopping. Controlled selective-repair experiments vary only which observed non-stop loss terms remain active, keeping examples, content and complete-stop losses, denominator, and optimization budget fixed. Retaining 820 high-non-stop-loss positions among 8,192 eligible positions improves flat ROUGE-L by .0694 over matched non-high-risk random support and by .0679 over conditional-token-difficulty support on 256 articles. Removing those positions produces the complementary loss of overlap. On 1,536 articles, high-loss retention remains .0571 above full-pool random support and .0007 below full support; no noninferiority margin was specified. Crossing stopping rules between the same removal endpoints reproduces a +.0675 overlap difference against a total +.0686, linking the support intervention to learned stopping. A training-selected global EOS bias recovers .0483 ROUGE-L but remains .0228 below donor stopping. Automatic source-support declines as overlap recovers. Additional controls delimit the explanation: matching initial scalar non-stop pressure does not reproduce the large repair; native removal and XSum do not establish a high-loss advantage; and the same selected supports have a smaller contrast from a native parent. The contribution is a controlled account of continuation supervision in the tested repair regime. Parameter-level mechanisms, factuality gains, and improvements over ordinary native training remain unestablished.

1 INTRODUCTION

Stopping is part of language generation, not merely a decoding option. A model can learn useful next-token preferences yet terminate before those preferences are expressed. Conversely, extending a response can increase reference overlap while exposing unsupported content. Analyses of the end-of-sequence decision and length extrapolation establish this distinction (Newman et al., 2020; Sountsov & Sarawagi, 2016); selective EOS supervision has also been studied in multimodal generation (Yue et al., 2024). The remaining question here is narrower: *which observed continuation terms actually change a learned stopping rule?*

An observed non-stop token provides two pieces of supervision: continuation occurred, and a particular token occurred conditional on continuation. Separating these terms permits an intervention on continuation supervision while every content target remains trained. The central finding is that, even with complete STOP supervision also fixed, the support of observed continuation terms can strongly change learned stopping. Content-target coverage alone therefore does not characterize the behavioral effect of a supervision mask.

Selective parents train conditional content and complete-stop losses without continuation supervision; native parents also train continuation loss. The repair recipe is fixed. Removing high-continuation-loss support H sharply reduces overlap; retaining H alone recovers most of the full-support difference. Matched non-high-risk random, full-pool random, and conditional-token-difficulty controls establish which tested alternatives reproduce this effect. Functional crosses of the resulting endpoints then test whether their learned stopping rules carry the behavioral difference.

The comparisons link the loss intervention to generated behavior (Table 1). Full-pool random partly repairs the selective endpoint; initial scalar-pressure matching does not reproduce the large recovery. Native removal, parent crossover, XSum and the weak second-model effect bound the result.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

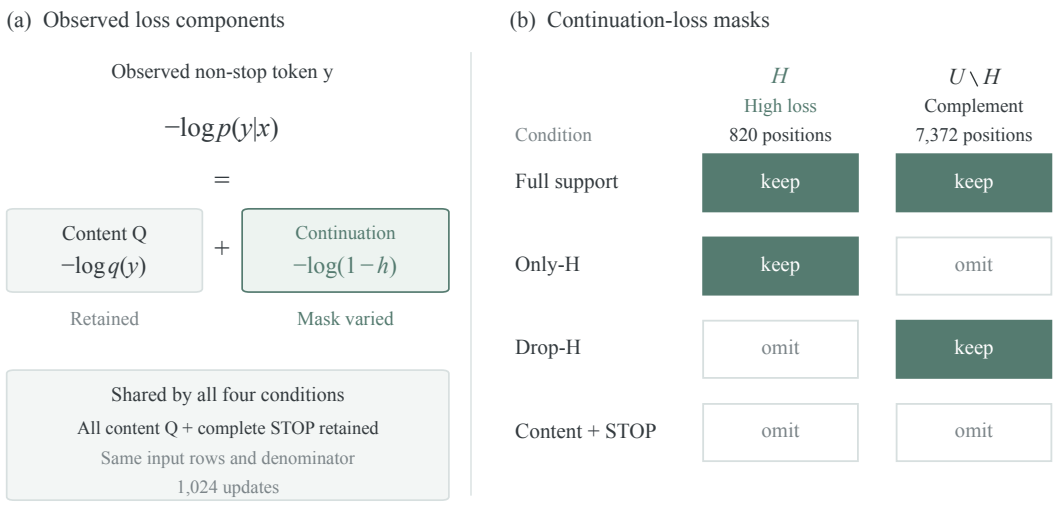

Figure 1: Observed losses and selective continuation supervision. (a) An observed non-stop token contributes the sum of content Q and continuation losses. Content and complete STOP supervision remain active in every arm; Content + STOP omits all continuation-loss terms. (b) Each row directly shows which continuation terms remain active. H contains the 820 positions with the largest initial continuation losses in the 8,192-position eligible set U; the complement $U \setminus H$ contains 7,372. Equal-width columns identify sets, not token order or size. "Omit" removes only the continuation term, preserving input rows, denominator and 1,024 updates; content parameters remain trainable. Functional decoder responses appear in Figure 3.

Table 1: Three distinct questions and their answers. All differences are raw flat ROUGE-L unless explicitly labeled otherwise. The 1,536-article panel is a separate larger confirmation.

Question	Controlled comparison	Finding and boundary
Which training terms matter?	Only H versus non-high-risk / full-pool random; same content and STOP losses	+.0675 / +.0571 on 1,536 articles. Not a 90% data or compute saving.
How does generation change?	Exchange stopping rules between the same removal endpoints	Gate transfer +.0675 versus total +.0686 on 1,536 articles. Not a natural mediation fraction.
Does more overlap mean better content?	Source-support, both forced lengths, and a native-objective baseline	Support decreases; content contrasts depend on length; native overlap is higher. No factuality claim.

The diagnostic separates continuation-event supervision from content-target coverage and evaluates stopping behavior, establishing neither a general selector nor a factuality improvement.

2 Related Work

Stopping, length, and generation. EOS decisions influence extrapolation (Newman et al., 2020), length bias can arise from locally normalized sequence models (Sountsov & Sarawagi, 2016), and sequence oversmoothing affects autoregressive behavior (Kulikov et al., 2022). The consistency of incomplete decoding and the degeneracy of common decoding procedures further separate a trained distribution from its generated text (Welleck et al., 2020a; Holtzman et al., 2020). Unlikelihood training intervenes on undesirable generation events (Welleck et al., 2020b); selective EOS supervision directly studies the EOS/content distinction in multimodal models (Yue et al., 2024). The present experiments instead connect a fixed, observed binary-loss support to learned stopping through matched removal, retention, and endpoint-crossing comparisons. The comparisons do not infer a natural termination hazard from externally truncated responses.

Selective learning and information access. Importance sampling exploits heterogeneous sample contributions (Katharopoulos & Fleuret, 2018). Rho-1 selects token cross-entropies using a reference and current-model excess-loss criterion (Lin et al., 2024). Here, all conditional content targets remain trained, and only the observed non-stop component is masked. The Q-hard control is a simple conditional-token-difficulty selector, not an implementation of Rho-1. Work on novel-token learning (Lin et al., 2021), truncation in language modeling (Ding et al., 2024), and complementary-label learning (Chou et al., 2020) motivates care about the signal that an objective actually supplies. A valid loss decomposition alone does not establish effective optimization. Likewise, prefix-distribution changes matter in sequential prediction (Ross et al., 2011); fixed-prefix diagnostics and free-running generation answer different questions.

Evaluation is a separate measurement problem. ROUGE measures reference overlap (Lin, 2004), while embedding-based metrics such as BERTScore measure a different similarity relation (Zhang et al., 2020). Neither is a count of correct facts. Factual-consistency studies document this gap (Maynez et al., 2020; Kryscinski et al., 2020; Pagnoni et al., 2021), and NLI-based metrics depend on segmentation and aggregation (Laban et al., 2022). AlignScore provides a useful automatic alignment measure (Zha et al., 2023); its value here remains a source-support proxy. Atomic factual precision, as operationalized by Min et al. (2023), requires an additional measurement procedure that these experiments do not validate. The distinction determines which quality claims the present comparisons support.

3 A CONTROLLED INTERVENTION ON CONTINUATION SUPPORT

3.1 SEPARATE THE OBSERVED EVENT FROM ITS TOKEN IDENTITY

Let $p_\theta(\cdot \mid x)$ be a finite-logit softmax distribution with nonempty stop set $\mathcal{S}$ and nonempty complement. Define

$$h_\theta(x) = \sum_{s \in \mathcal{S}} p_\theta(s \mid x), \qquad q_\theta(u \mid x) = \frac{p_\theta(u \mid x)}{1 - h_\theta(x)}, \quad u \notin \mathcal{S}. \tag{1}$$

For an observed non-stop token, the ordinary loss decomposes as

$$-\log p_\theta(u \mid x) = \underbrace{-\log(1 - h_\theta(x))}_{\ell_{\mathrm{cont}}(x)} + \underbrace{-\log q_\theta(u \mid x)}_{\ell_Q(x,u)}. \tag{2}$$

Here ℓ_{cont} is continuation loss and ℓ_Q is conditional-content loss. The intervention varies the support of the continuation term. Continuation supervision applies only where continuation is observed; complete STOP supervision applies only where the terminal event is observed. A truncated boundary is not silently relabeled as a natural stop.

For fixed inherited endpoint r, batch B, and continuation support A, the objective is

$$\mathcal{L}_A(B) = \frac{Q_B + S_B + \sum_{i \in A \cap \mathrm{eligible}(B)} \ell_{\mathrm{cont}}(x_i)}{D_B}. \tag{3}$$

Q_B, complete-stop contribution S_B, original denominator D_B, and all input rows are unchanged across arms. In particular, the denominator is not the number of active continuation terms. Full-network optimization can still change q_θ; unchanged *content supervision* does not mean frozen *content parameters*.

At the free-logit level, a common shift b of stop logits leaves q and ℓ_Q unchanged. With $a = \mathrm{LSE}(z_{\mathrm{STOP}}) - \mathrm{LSE}(z_{\mathrm{nonstop}})$, $h = \sigma(a)$ and $\ell_{\mathrm{cont}} = \mathrm{softplus}(a)$, so $\partial \ell_{\mathrm{cont}} / \partial b = h$ and $\|\nabla_z \ell_{\mathrm{cont}}\|_1 = 2h$. These elementary identities explain why matching content targets or continuation counts does not match continuation pressure. These identities imply neither independent prefix shifts inside a Transformer nor a parameter-level mechanism; raw-greedy stopping also depends on the largest competing token logit.

3.2 SELECT ONCE, THEN RETAIN OR REMOVE

The primary setting uses SmolLM2-1.7B-Instruct (Ben Allal et al., 2025) and two fixed inherited selective endpoints, trained with fixed within-block row permutations from seeds 2 and 3 (Appendix A). Each endpoint receives the same 16,384-row continuation tape for 1,024 updates, with

effective batch size 16. The set U contains each tape's 8,192 eligible last-visible-prefix continuation positions. At the inherited endpoint, H is the top $\lceil .1 \times 8192 \rceil = 820$ positions by ℓ_{cont}, with fixed-ID ties. A random set R excludes H and matches its counts in 16 strata formed by quartiles of the visible-target index $j = \max(0, n-1)$ and native prompt token length, where n counts visible assistant tokens. Thus R is a matched non-high-risk control, not an independent uniform subset of all eligible positions.

An additional R_{pool} samples uniformly without replacement from each *entire* stratum, retaining the same quotas and 820 positions. The single fixed draws overlap H in 80/78 positions (orders 2/3), against stratified expectations 86.068/85.797; no overlap-based redraw occurs.

For each order, the four supports $\varnothing, H, U \setminus H, U$ complete the factorial in Figure 1. The corresponding endpoints are Content + STOP, only-H, drop-H, and full support. Only-R and drop-R provide matched controls. A Q-hard set selects the same stratum quotas by ℓ_Q on allowed training targets. Its training intervention remains the same continuation term in Equation 3, without selecting or reweighting the Q loss. All 8,192 positions have a visible, permitted target for this selector.

Training preserves the inherited FP32 SDPA kernel and fresh AdamW recipe: learning rate 10^{-5}, betas $(.9, .999)$, epsilon 10^{-8}, weight decay .01 applied as decoupled decay, and gradient clipping at 1. Optimizer state evolves continuously between monitors. The exact implementation matters, including the distinction between optimizer states and weight-decay conventions (Kingma & Ba, 2015; Loshchilov & Hutter, 2019). No arm receives a longer horizon because its effect is weak. Appendix A records the design and provenance.

3.3 Evaluation and estimands

The task is article-conditioned generation in the CNN/DailyMail summarization setting (See et al., 2017). The central confirmation contains 256 fixed source-selected articles and both training orders. A separately fixed larger panel contains 1,536 articles and the same endpoints. The larger panel was not used to alter the 256-panel primary or tune the masks. Available local exposure records delimit selection; neither panel is claimed globally absent from all past experiments or pretraining. Evaluation articles are not new training replicates.

Let $F(A)$ denote a metric evaluated after training with continuation support A. The complement estimands are

$$\Delta_H^{\text{add}} = F(H) - F(\varnothing), \qquad \Delta_H^{\text{remove}} = F(U) - F(U \setminus H),$$
$$J_H = \Delta_H^{\text{remove}} - \Delta_H^{\text{add}}. \tag{4}$$

The removal contrast is full − drop-H: positive ROUGE-L means a score loss upon removing H, not the after − before change. These are trained-endpoint differences, not a linearity assumption. The retention primary is only-H − only-R flat ROUGE-L. Other metrics, J_H, J_R, and selector contrasts retain their secondary roles. The original removal comparison remains a separate experiment. All displayed overlap values use the raw $[0, 1]$ scale.

Paired article bootstrap resampling keeps all arms and both orders together (4,096 draws). Reported intervals therefore quantify article variation conditional on the fixed trained endpoints. Order-specific effects accompany aggregate means. The intervals do not cover a population of independently trained models; sources of training variance require a different design (Dror et al., 2018; Bouthillier et al., 2021). No practical noninferiority margin was fixed, so similar means or intervals containing zero do not establish equivalence or sufficiency. The later full-pool, offset, native-removal, parent-crossover and XSum contrasts are fixed amendments with descriptive 95% intervals, not a retrospectively corrected joint confirmatory family; the original primary and interval levels remain.

4 Concentrated continuation support recovers overlap

4.1 Removal and retention tell a consistent, bounded story

On 256 articles, full support reaches .288241 flat ROUGE-L, compared with .217742 for Content + STOP and .217230 for drop-H. Only-H reaches .286564; only-R reaches .217178. Thus keeping H

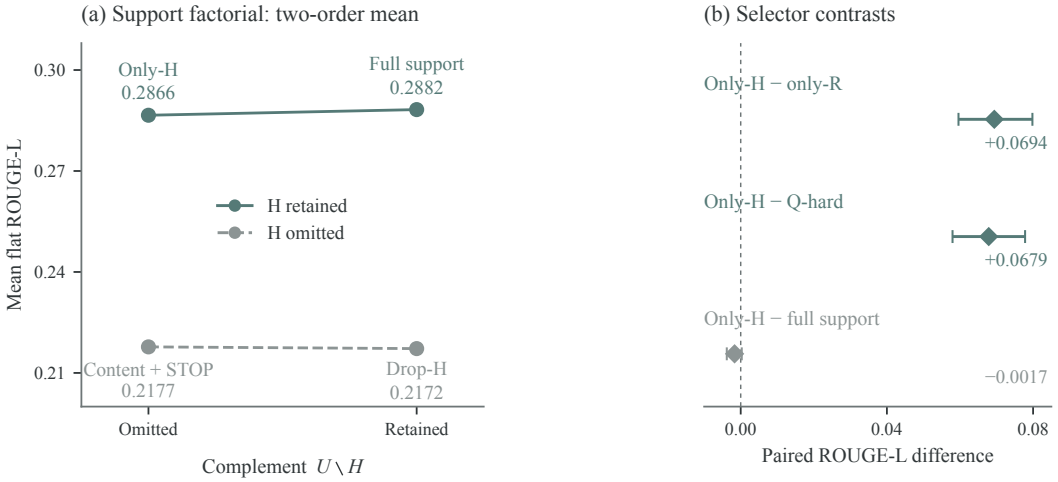

Figure 2: Complementary support and selector controls on 256 shared articles. (a) Circles average the two fixed training orders (seeds 2 and 3); Table 7 retains order-specific values. Solid and dashed segments vary complement $U \setminus H$ support with H retained and omitted, respectively, at fixed training time. (b) Diamonds are paired means and lines are 95% article-paired intervals conditional on the two fixed training orders. Only-H, only-R and Q-hard use equally many continuation positions; full support is a separate reference. No absolute-score intervals or independent training replications are inferred.

Table 2: Retention and signal-control evidence. All intervals are 95% paired article intervals. Later signal controls and XSum are descriptive amendments; neither panel enlargement nor a new control replaces the original primary.

Panel	Contrast / metric	Difference	Interval
256	Only-H − only-R / ROUGE-L	+.069385	[.059588,.079814]
	Only-H − Q-hard / ROUGE-L	+.067867	[.057958,.077811]
	Only-H − full / ROUGE-L	−.001678	[−.003808,.000285]
	Only-H − R_{pool} / ROUGE-L	+.058827	[.049060,.068722]
1,536	Only-H − only-R / ROUGE-L	+.067477	[.063354,.071796]
	Only-H − full / ROUGE-L	−.000708	[−.001676,.000270]
	Only-H − R_{pool} / ROUGE-L	+.057129	[.053339,.060894]
	R_{offset} − R / ROUGE-L	+.000542	[−.000377,.001530]
	Native dropR − dropH / ROUGE-L	+.000322	[−.000561,.001234]
XSum 256	H − R_{pool} / ROUGE-L	−.000516	[−.002845,.001744]

alone produces a large recovery even though the remaining continuation terms are absent (Figure 2). Only-H−only-R is +.069385, with 95% interval [.059588,.079814]. The two order effects, +.068480 and +.070291, agree in sign and scale. Only-H − Q-hard is +.067867 [.057958,.077811]; Q-hard reaches .218696.

The addition and removal effects are +.068822 and +.071011, respectively. Their difference J_H is +.002190 [−.000033,.004526], while J_R is +.000031 [−.002214,.002242]. The large removal effect therefore does not require a large estimated interaction with the complementary support in this panel. However, the intervals do not prove additivity. Only-H − full is −.001678 [−.003808,.000285], leaving both a small deficit and near-zero difference compatible with the observations. Calling the subset "sufficient" would add an untested tolerance claim.

The larger panel preserves the result: only-H − only-R is +.067477 [.063354,.071796], with order effects +.063635 and +.071319. Mean content lengths are 52.463 and 20.760 tokens, respectively; full support averages 52.507. All these selective endpoints reach EOS below the cap with their original decoders. Equal EOS incidence plainly does not imply equal stopping time. The larger panel does not include Content + STOP or Q-hard, and therefore supplies neither a new complement in-

Table 3: The selector control separates continuation difficulty from Q difficulty, but not from initial signal dose. Profiles use training positions only. Scalar mass includes occurrence and original-denominator weights, not parameter-gradient norms or accumulated Adam updates.

Quantity	Order 2	Order 3
$\lvert H \cap H_Q \rvert$	41 / 820	40 / 820
Jaccard(H, H_Q)	.025641	.025000
Spearman$(\ell_{\mathrm{cont}}, \ell_Q)$	.024515	.037711
continuation-loss mass on H	4.082100	4.091823
continuation-loss mass on R	$7.0664 \cdot 10^{-10}$	$8.7626 \cdot 10^{-10}$
Scalar derivative mass on H	.658559	.652699
Scalar derivative mass on $U \setminus H$	$6.9379 \cdot 10^{-9}$	$7.9977 \cdot 10^{-9}$
Scalar derivative mass on R_{pool}	.061180	.060484
Offset online / frozen-parent scalar sum	.668832	.657766

teraction nor a larger-panel Q-hard result. The additional R_{pool} control improves over R by .010348 [.008758,.011987] yet remains below H in both orders (.057232/.057026). Mean length rises to 24.529 tokens, versus 52.463 for H. Thus ordinary full-pool random partly works; these two draws neither establish all-random failure nor identify an optimal selector.

4.2 Recovery is not general quality improvement

The same only-H versus only-R comparison reduces automatic source-support: $-.028322$ [$-.043251$,$-.013996$] on 256 articles and $-.026555$ [$-.033843$,$-.019376$] on 1,536. More overlap and more generated content do not jointly imply better-supported claims. Only-H closely tracks full support on these metrics, but neither metric supports preservation of every article's wording or stopping index. On the larger panel, only-H and full have exactly matching token responses on 1,309/1,536 and 1,252/1,536 articles by order; aggregate proximity is weaker than behavioral identity.

The selected set comprises roughly 10% of *eligible continuation positions*, not 10% of all training tokens, examples, or supervision. The full Q computation, all rows, and complete STOP terms remain active. The comparison demonstrates concentrated corrective influence under a fixed objective and optimization budget, not a 90% training-cost reduction or minimal-support theorem.

5 Continuation difficulty differs from token difficulty

At a fixed observed prefix, ℓ_{cont} measures disagreement with the event "continue," whereas ℓ_Q measures difficulty predicting the observed token given continuation. The training profiles separate these notions before evaluation. The high-continuation-loss and Q-hard sets overlap in only 41 and 40 of their 820 positions in orders 2 and 3; their Jaccard similarities are .02564 and .02500. Spearman correlations between the two scores over all eligible positions are .02451 and .03771. Their distinction is therefore empirical, not just notational, in these inherited endpoints.

With the same stratum quotas, input rows and continuation loss, Q-hard does not reproduce the recovery obtained by H. Thus equally many difficult conditional-content positions do not suffice under this recipe. The comparison neither rejects selective language modeling in general nor establishes a continuation-loss selector's advantage over excess-loss selection on unrelated tasks.

Initial scalar matching is not trajectory matching. The initial derivative mass on H is about 10^8 times that on its complement (Table 3), leaving signal magnitude as a competing explanation. A fixed diagnostic surrogate applies $\lambda \,\mathrm{softplus}(a_j + d_j)$ at the existing R positions. Training-only offsets and weights match the initial aggregate common-stop-shift derivative to H, retaining Q, complete STOP, inference logits, D_B and 1,024 updates. This control changes the training objective without matching parameter gradients or subsequent optimization.

Offset $-$ R is $-.000236$ [$-.002343$,.001781] on 256 articles and remains inconclusive on 1,536 (Table 2); H exceeds offset by .066935 [.062615,.070959] on the latter panel. The specified initial

match does not reproduce the large repair. The offset control's online scalar sums are roughly two-thirds of the frozen-parent sums, while fixed-prefix first-stop indices remain unchanged. These observations do not isolate clipping, parameter direction, or cumulative dose as the cause. Appendix A preserves the surrogate and time-course diagnostics. Exhaustive training selection of a global EOS bias recovers 4.83 ROUGE-L points, but donor stopping remains 2.28 points better (Appendix C.2). Thus scalar correction is useful without reproducing the donor rule.

6 THE SAME REMOVAL ENDPOINTS CONNECT TRAINING TO STOPPING

6.1 CROSS STOPPING DECISIONS, THEN CONTROL LENGTH

Post-training crosses use the same trained drop-H and drop-R endpoints. Let $D(a, b)$ take model a's raw-greedy stop/continue decision at the current evolving prefix. If continuation is selected, model b supplies its highest probability non-stop token. Stop-token sets and native tie rules are unchanged. This is a deterministic intervention on fixed functions, requiring two model forwards per step. Native diagonal recovery is checked before interpreting off-diagonals. Because changed content also changes later prefixes, the cross is not a causal mediation analysis of training.

Write $R = \text{drop-}R$ and $H' = \text{drop-}H$ for this subsection only. Their endpoint gap, distinct from Δ_H^{remove}, decomposes algebraically as

$$F(R) - F(H') = \underbrace{F(D(R, H')) - F(H')}_{\text{gate transfer}} + \underbrace{F(R) - F(D(R, H'))}_{\text{content at R gate}}. \tag{5}$$

The two off-diagonals are generated, rather than extrapolated from an older pair with and without last-visible-prefix continuation supervision. On 256 articles, the overlap gate transfer is +.071103, with the prespecified 97.5% interval [.059825,.082965], against a total removal difference of +.071544. Content at the R gate is +.000441 [−.002196,.003268] at the same confidence level.

On 1,536 articles (Figure 3), gate transfer is +.067504 [.063377,.071801], against total +.068576 [.064519,.072920]. Content at the R gate is +.001072 [.000068,.002115]. Thus a small positive overlap content contrast can coexist with a much larger stopping contrast. This descriptive ROUGE-L contrast is distinct from the inconclusive 128-article source-support primary at matched long lengths (Section 7); the 256-article gate-content contrast was secondary. Reporting the absolute effects also avoids presenting a point-estimate recovery ratio as an identified mediation percentage.

Both-length crosses constrain the interpretation further. The drop-H content model is continued to drop-R's native length, and drop-R is cut to drop-H's native length. No reference length is used. On 1,536 articles, the overlap content contrasts are +.001120 [.000149,.002127] at the longer length and +.000270 [−.000394,.000954] at the shorter length. In contrast, source-support is +.005403 [.002646,.008005] at the longer length but −.005588 [−.008579,−.002618] at the shorter length. Content quality cannot be summarized by a length-invariant scalar improvement from these comparisons.

The long-length contrast compares original drop-R with extended drop-H; the short-length contrast compares truncated drop-R with original drop-H. Each contrast matches token length, but forced termination can interrupt a sentence. Section 7 addresses completion without replacing the exact-token comparison.

6.2 A STRONG NATIVE COMPARATOR LIMITS PRACTICAL CLAIMS

A native-objective comparator starts from the native-objective parent M and receives the same additional 1,024-update tape. Its different warm objective and weights make this an exposure-matched contextual baseline. On 1,536 articles, full support − native M is −.006326 [−.009964,−.002746] in overlap, with order effects −.012369/−.000284. Native M averages 95.784 content tokens versus 52.507 for full support and hits the cap in .749% of outputs; all capped rows remain. Full support has higher source-support by .011562 [.006544,.016581] but lower highlight support by .121140. Repair of the selective endpoint therefore establishes neither better ordinary native training nor a recommended serving policy. Two-forward donor decoding also incurs diagnostic inference overhead.

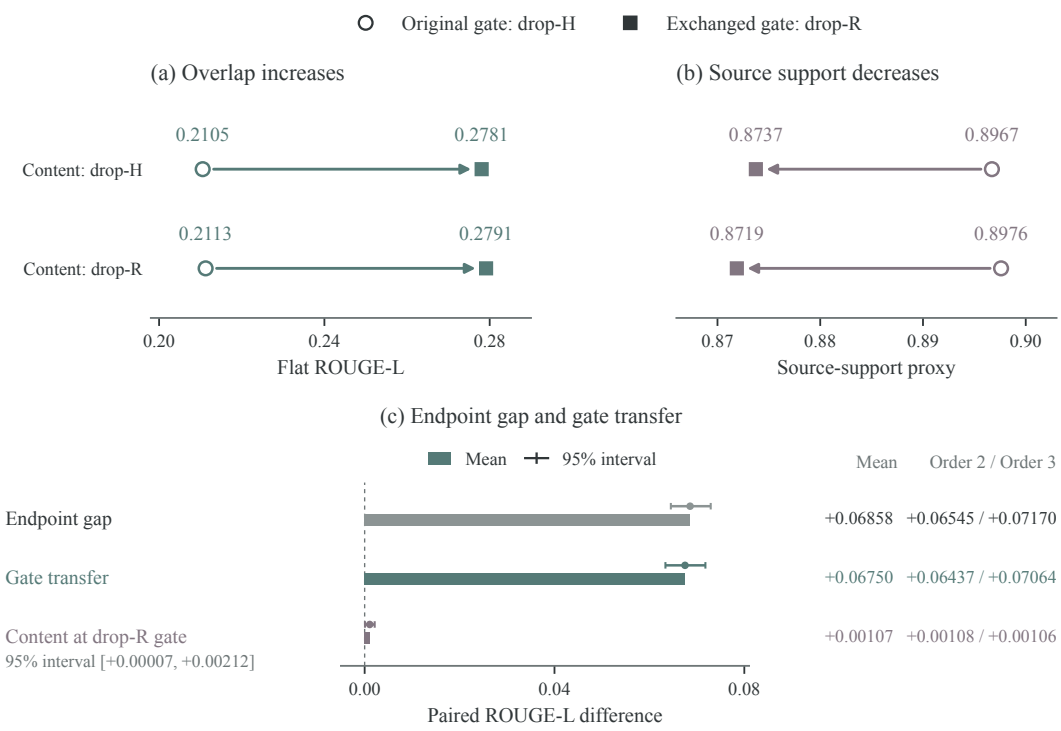

Figure 3: Crossed decoder responses on 1,536 fixed articles. (a,b) Each row fixes the content donor. Arrows exchange the drop-H gate (open circle) for the drop-R gate (filled square): overlap rises while automatic source support falls. Endpoints are absolute means, not confidence limits; neither metric is factual precision. (c) Gate transfer nearly matches the drop-R − drop-H endpoint gap. Bars show paired means; offset capped lines show descriptive 95% article intervals conditional on two fixed orders. Order-specific effects are printed separately. Native diagonals reproduce their corresponding native outputs; no mediation percentage is identified. Tables 11–12 retain complete contrasts, and Table 4 retains forced-length and native-M comparisons.

Table 4: Fixed conditions on the same 1,536 articles, including full-pool random. Both forced-length directions and the stronger native comparator remain visible alongside retention. Support is an automatic source-support proxy, not factual precision. Mean lengths are native unless marked as forced.

Output	ROUGE-L	Content tokens	Support
Only-H	.278593	52.463	.873547
Only-R	.211116	20.760	.900102
R_{pool}	.221464	24.529	.895028
Full support	.279301	52.507	.873067
Drop-H	.210549	20.723	.896708
Drop-R	.279126	52.688	.871884
$D(\text{dropR}, \text{dropH})$	.278054	52.477	.873735
$D(\text{dropH}, \text{dropR})$	.211258	20.798	.897600
Drop-H at drop-R length (forced)	.278005	52.688	.866482
Drop-R at drop-H length (forced)	.210819	20.723	.891120
Native M	.285627	95.784	.861505

Same-M removal provides a matched test. Training-only H_M and matched non-high-risk R_M use the original 8,192 eligible positions; removal preserves all other native visible continuation terms. The same-parent drop-R − drop-H contrast is −.000502 [−.002674,.001663] on 256 articles and +.000322 on 1,536 (Table 2). Exact parent weights, data order, numerical kernel and optimizer set-

tings permit reuse of native Full; Full − drop-H is .000267 [−.000716,.001278] on 1,536. Neither large H-specific native damage nor equivalence is established. Both parent state and continuation background differ from selective repair. Saved S/M profiles have similar rankings but lower native stop mass (Appendix A). The crossover changes the parent with S-selected supports and selective continuation background fixed. On 256 articles, $H_S - R_{\text{pool},S}$ is +.022369 [.015979,.029293] from M, smaller than from S by .036458 [.027444,.045404]. Yet M's H_S continuation − its uncontinued parent is only +.006803 [−.002349,.015981]. Relative selection advantage does not imply improvement over the parent. This exposed comparison bounds regime dependence without identifying a parameter mechanism or testing full-native supervision (Appendix B).

7 Measurement boundaries and scope

Support scores are not a fact inventory. Source-support uses fixed AlignScore-base sentence/chunk aggregation; the highlight proxy reverses the query direction using literal reference-highlight lines. Neither measures all source facts, atomic factual precision or factual coverage. A local fact-level evaluator failed prespecified tests, including elementary source-support controls: a measurement failure, not a scientific null about generated claims. No human factual annotation is reported; an unvalidated replacement judge cannot fill that gap.

Earlier positive content estimates did not confirm. Both order pairs' earlier long-continuation source-support primaries on the same 128 articles are inconclusive: +.010868 [−.004528,.025731] and +.004954 [−.011998,.021046] (97.5%). Secondary interactions do not reverse these outcomes. Saved-output sentence trimming also gives inconclusive short/long effects (+.000238/−.000641), conditional on 114/125 paired-complete articles. Trimming changes length, precluding an equal-token primary; Appendix D reports complete values and missingness.

Generalization and practical interpretation. The strongest evidence concerns one SmolLM2 repair regime with two inherited orders. Four historical orders share a pretrained family and source, without independent pretraining. The historical Qwen2.5-3B long comparison (Qwen, 2024) gives +.000036 [−.005776,.005905] for last-visible-prefix continuation, without replicating the large Smol effect across families. XSum (Narayan et al., 2018) supplies a source boundary with two fresh native parents, 2,048-update warm starts, training-only selection and 1,024-update Full/H/R_{pool} continuations. On 256 source-selected articles, $H - R_{\text{pool}}$ is −.000516 [−.002845,.001744]; H − Full is −.001384 [−.003585,.000605]. All arms reach EOS without capping, averaging about 25 content tokens. Native-complete eligibility includes all observed preterminal continuation positions, unlike CNN's selective/censored rule. Source-only length filtering, one pretrained ancestry and unknown pretraining exposure limit transport. XSum establishes no H advantage. No selector optimality, minimum-support or factuality claim follows.

8 Conclusion

Selective repair changes learned stopping when retained continuation terms change, despite fixed supervision for all content targets and complete STOP. Complementary removal and retention establish the support effect; random and Q-hard controls delimit selection, and crossing the same endpoints links the effect to stopping. Full-pool random partly repairs the model; initial scalar matching does not reproduce the large recovery. Global EOS correction recovers much of the overlap gap but remains below donor stopping. Native, crossover and XSum results bound the findings, while automatic source-support declines as overlap returns. The evidence supports a component-specific behavioral effect in this regime; a general stopping method, complete optimization mechanism and factuality improvement remain unestablished.

AI USE STATEMENT

Generative AI tools assisted the design, implementation and analysis of training and decoding experiments, mathematical formulation and checking, literature search and reference verification, figure preparation, and manuscript drafting and editing. The authors are responsible for the final content.

ETHICS STATEMENT

The study uses existing language models and article-conditioned generation. Longer generations can expose more unsupported content; overlap improvements should not be interpreted as factuality or deployment-safety improvements. Source data and model licenses remain applicable. Automatic source-support scores are proxies, and no validated human or AI-generated factual annotations are claimed.

REPRODUCIBILITY STATEMENT

Appendices A–C describe training, the reported contrasts and decoder validation. Appendix D documents measurement sensitivity; Appendix D.1 summarizes the scope of the reported evidence. Effect estimates exclude unfinished experiments and measurements from the unvalidated factuality evaluator; valid adverse results and capped outputs remain included.

REFERENCES

Loubna Ben Allal, Anton Lozhkov, Elie Bakouch, Gabriel Martín Blázquez, Guilherme Penedo, Lewis Tunstall, Andrés Marafioti, Hynek Kydlíček, Agustín Piqueres Lajarín, Vaibhav Srivastav, et al. SmolLM2: When smol goes big—data-centric training of a small language model. *arXiv preprint arXiv:2502.02737*, 2025. URL https://arxiv.org/abs/2502.02737.

Xavier Bouthillier, Pierre Delaunay, Mirko Bronzi, Assya Trofimov, Brennan Nichyporuk, Justin Szeto, Nazanin Mohammadi Sepahvand, Edward Raff, Kanika Madan, Vikram Voleti, Samira Ebrahimi Kahou, Vincent Michalski, Tal Arbel, Chris Pal, Gael Varoquaux, and Pascal Vincent. Accounting for variance in machine learning benchmarks. In *Proceedings of Machine Learning and Systems*, 2021. URL https://proceedings.mlsys.org/paper_files/paper/2021/hash/0184b0cd3cfb185989f858a1d9f5c1eb-Abstract.html.

Yu-Ting Chou, Gang Niu, Hsuan-Tien Lin, and Masashi Sugiyama. Unbiased risk estimators can mislead: A case study of learning with complementary labels. In *Proceedings of ICML*, 2020. URL https://proceedings.mlr.press/v119/chou20a.html.

Hantian Ding, Zijian Wang, Giovanni Paolini, Varun Kumar, Anoop Deoras, Dan Roth, and Stefano Soatto. Fewer truncations improve language modeling. In *Proceedings of ICML*, 2024. URL https://proceedings.mlr.press/v235/ding24f.html.

Rotem Dror, Gili Baumer, Segev Shlomov, and Roi Reichart. The hitchhiker's guide to testing statistical significance in natural language processing. In *Proceedings of ACL*, 2018. URL https://aclanthology.org/P18-1128/.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. The curious case of neural text degeneration. In *Proceedings of ICLR*, 2020. URL https://arxiv.org/abs/1904.09751.

Angelos Katharopoulos and François Fleuret. Not all samples are created equal: Deep learning with importance sampling. In *Proceedings of ICML*, 2018. URL https://proceedings.mlr.press/v80/katharopoulos18a.html.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of ICLR*, 2015. URL https://arxiv.org/abs/1412.6980.

Wojciech Kryscinski, Bryan McCann, Caiming Xiong, and Richard Socher. Evaluating the factual consistency of abstractive text summarization. In *Proceedings of EMNLP*, 2020. URL https://aclanthology.org/2020.emnlp-main.750/.

Ilia Kulikov, Maksim Eremeev, and Kyunghyun Cho. Characterizing and addressing the issue of oversmoothing in neural autoregressive sequence modeling. In *Proceedings of AACL*, 2022. URL https://aclanthology.org/2022.aacl-main.82/.

Philippe Laban, Tobias Schnabel, Paul N. Bennett, and Marti A. Hearst. SummaC: Re-visiting NLI-based models for inconsistency detection in summarization. *Transactions of the Association for Computational Linguistics*, 10:163–177, 2022. URL https://aclanthology.org/2022.tacl-1.10/.

Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, 2004. URL https://aclanthology.org/W04-1013/.

Xiang Lin, Simeng Han, and Shafiq Joty. Straight to the gradient: Learning to use novel tokens for neural text generation. In *Proceedings of ICML*, 2021. URL https://proceedings.mlr.press/v139/lin21b.html.

Zhenghao Lin, Zhibin Gou, Yeyun Gong, Xiao Liu, Yelong Shen, Ruochen Xu, Chen Lin, Yujiu Yang, Jian Jiao, Nan Duan, and Weizhu Chen. Not all tokens are what you need for pretraining. In *Advances in Neural Information Processing Systems*, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/3322a9a72a1707de14badd5e552ff466-Abstract-Conference.html.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *Proceedings of ICLR*, 2019. URL https://arxiv.org/abs/1711.05101.

Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. On faithfulness and factuality in abstractive summarization. In *Proceedings of ACL*, 2020. URL https://aclanthology.org/2020.acl-main.173/.

Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Wei Koh, Mohit Iyyer, Luke Zettlemoyer, and Hannaneh Hajishirzi. FActScore: Fine-grained atomic evaluation of factual precision in long form text generation. In *Proceedings of EMNLP*, 2023. URL https://aclanthology.org/2023.emnlp-main.741/.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1797–1807, 2018. doi: 10.18653/v1/D18-1206. URL https://aclanthology.org/D18-1206/.

Benjamin Newman, John Hewitt, Percy Liang, and Christopher D. Manning. The EOS decision and length extrapolation. In *Proceedings of BlackboxNLP*, 2020. URL https://aclanthology.org/2020.blackboxnlp-1.26/.

Artidoro Pagnoni, Vidhisha Balachandran, and Yulia Tsvetkov. Understanding factuality in abstractive summarization with FRANK: A benchmark for factuality metrics. In *Proceedings of NAACL*, 2021. URL https://aclanthology.org/2021.naacl-main.383/.

Qwen. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2024. URL https://arxiv.org/abs/2412.15115.

Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of AISTATS*, 2011. URL https://proceedings.mlr.press/v15/ross11a.html.

Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. In *Proceedings of ACL*, 2017. URL https://aclanthology.org/P17-1099/.

Pavel Sountsov and Sunita Sarawagi. Length bias in encoder decoder models and a case for global conditioning. In *Proceedings of EMNLP*, 2016. URL https://aclanthology.org/D16-1158/.

Sean Welleck, Ilia Kulikov, Jaedeok Kim, Richard Yuanzhe Pang, and Kyunghyun Cho. Consistency of a recurrent language model with respect to incomplete decoding. In *Proceedings of EMNLP*, 2020a. URL https://aclanthology.org/2020.emnlp-main.448/.

Sean Welleck, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. Neural text generation with unlikelihood training. In *Proceedings of ICLR*, 2020b. URL https://arxiv.org/abs/1908.04319.

Zihao Yue, Liang Zhang, and Qin Jin. Less is more: Mitigating multimodal hallucination from an EOS decision perspective. In *Proceedings of ACL*, 2024. URL https://aclanthology.org/2024.acl-long.633/.

Yuheng Zha, Yichi Yang, Ruichen Li, and Zhiting Hu. AlignScore: Evaluating factual consistency with a unified alignment function. In *Proceedings of ACL*, 2023. URL https://aclanthology.org/2023.acl-long.634/.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. BERTScore: Evaluating text generation with BERT. In *Proceedings of ICLR*, 2020. URL https://arxiv.org/abs/1904.09675.

A TRAINING DESIGN AND OBSERVED-LABEL RULES

Orders 2 and 3 are fixed row permutations generated with seeds 2 and 3. Each of two 16,384-row source blocks is shuffled separately with the same within-order seed, preserving block boundaries. Each order trains its own selective and native-objective parents from the same pretrained weights and 32,768 rows. Removal and retention reuse the corresponding parent's first block in the same order for 1,024 updates, with fresh empty Adam state. The earlier 3,072-update continuation appends the first block to the two-block sequence. The order seed also initializes training RNG. Shared pretrained ancestry and data preclude interpreting article-paired intervals as uncertainty over independent pretrained models.

Table 5: Training conditions for the short, matched support comparisons. "Kept" means the same supervision and coefficient, not frozen learned parameters. Native M is a contextual comparator with a different warm-start objective and weights but shared pretrained ancestry; its continuation support includes all visible non-stop positions.

Endpoint	continuation support	continuation positions	Q / complete STOP	Updates
Content + STOP	$\varnothing$	0	kept / kept	1,024
Only-H	H	820	kept / kept	1,024
Only-R	R	820	kept / kept	1,024
Q-hard-only	H_Q	820	kept / kept	1,024
Drop-H	$U \setminus H$	7,372	kept / kept	1,024
Drop-R	$U \setminus R$	7,372	kept / kept	1,024
Full support	U	8,192	kept / kept	1,024
Native M	all visible	not restricted to U	native objective	1,024

Fixed optimizer and data exposure. The short comparisons retain the 16,384-row tape, batch size 16, FP32 SDPA, fresh AdamW with foreach and fused implementations disabled, learning rate 10^{-5}, betas $(.9, .999)$, epsilon 10^{-8}, decoupled weight decay .01, clipping norm 1, and the original denominator. Each trajectory carries its actual optimizer state through the fixed update sequence. Monitoring at 0/32/128/256/1,024 does not create additional seeds or reset the optimizer. Matching the active continuation count without matching the original denominator would define a different experiment.

Selection information. H is computed once at the inherited endpoint from observed continuation losses. The 16 matching strata use the zero-based retained target index $j = \max(0, n-1)$ and prompt token length p. The corresponding causal-logit position is $p + j - 1$, distinct from the censored boundary index n. The native prompt includes the chat instruction, source article and assistant-generation prefix, excluding visible assistant tokens. Sorted eligible values use cut indices $\lceil Nq \rceil - 1$ for $q \in \{.25, .5, .75\}$; a bin counts cuts strictly below the value, keeping ties in the lower bin. Actual cuts are $(25, 32, 40)$ for j and $(613, 859, 1166)$ for p in both orders. Cartesian products define the strata without forcing equal row counts. R is disjoint from H, with fixed counts within each stratum. Q-hard uses the same quotas and the conditional Q loss of allowed visible targets. No evaluation answer or reference highlight enters these masks. Q-hard may overlap H; overlap is measured, not forcibly removed.

Observed supervision. The decomposition in Equation 2 supplies a binary continuation term when continuation is observed and a conditional token term when its target identity is visible. A complete terminal observation supplies STOP supervision. Externally truncated or identity-hidden boundaries do not reveal a natural EOS target by assumption. The experiments vary which observed continuation terms enter training without inferring unobserved labels from a reference. This identifies a comparison within the specified observation contract, not a survey of natural censoring mechanisms.

Historical context retained as a boundary. The earlier long Smol comparison used 3,072 additional updates. Adding continuation supervision at the last visible prefix gave +.066926 [.061910,.072028]; adding the term at the censored boundary gave −.000164 [−.002113,.001795] (95% and 99% intervals, respectively), both versus content and STOP alone. These effects motivated

support-specific intervention and are not pooled with the short trajectories. The earlier native comparator exceeds the last-visible-prefix condition by .005071 [.000940,.009351] at 97.5%. A Qwen long comparison is near zero.

Parents, eligible indices and denominators. The selective (S) and native-objective (M) parents share pretrained Smol ancestry and the original 32,768-row, 2,048-update warm tape. S trains conditional visible-token Q and complete STOP without visible continuation terms; M adds all visible continuation terms. The continuation uses the same first 16,384 rows, comprising 8,192 complete STOP rows and 8,192 non-stop rows. For visible length n, the eligible non-stop index is $j = \max(0, n-1)$ and its causal logit is at prompt length $+j-1$; all current non-stop rows have positive visible length. A row contributes $n+1$ to D_B, summed over all 16 logical rows. Visible target j uses logit prompt length $+j-1$; STOP id is 2. Native chat prompts and the original visible prefixes are preserved, with no hidden suffix supplied. The source was fixed by retokenizing the earlier Qwen-visible raw prefixes with Smol, not by recensoring half of a different vocabulary's tokens.

Selection and exposure of the fixed article panel. The 256-article panel is drawn from prequalified CNN/DailyMail test components. The inherited cross-split census links records sharing an ID, whitespace-normalized article or reference-highlight identity. Components touching training, validation or recorded prior evaluation exposure are excluded, with the lexicographically first test ID representing each component. Nonempty article/reference fields, a valid native chat prompt, prompt length plus 512 at most 8192, and absence of tokenizer special-token literals in the article determine eligibility.

After exclusion of recorded local source/evaluation components and exact or whitespace-normalized article matches, 4,144 components remain. A single fixed salted SHA256 ordering of article IDs selects the first 256, with ID tie-breaking and final ID sorting. The draw precedes reading selected highlights and model outputs; the inherited component census had already used reference identity. Prompt revalidation permits no drop or refill. These checks establish recorded-local separation, not absence of semantic near duplicates, all historical use or pretraining exposure.

Full-pool, offset and native controls. Full-pool random support is selected once with a fixed hash salt. Stratum quota $k_s = |H \cap U_s|$ gives expected overlap $\sum_s k_s |H \cap U_s|/|U_s|$, not an assumed 82 positions. Existing R remains H-excluded. Offset pairs sorted H/R identities within each stratum, using $d_j = a_{\theta_0}(x_{\pi(j)}) - a_{\theta_0}(x_j)$ from saved stable log-sum-exp margins. Training-only $\lambda = 1.008032/1.002574$ matches the initial occurrence- and denominator-weighted aggregate common-stop-shift derivative; the allowed range $[.5, 2]$ is not clamped. Q, complete STOP and inference logits receive no offset. This diagnostic surrogate differs from ordinary event likelihood and inference-time EOS bias. Both selective arms receive fresh AdamW and 1,024 updates. Native H_M/R_M are independently selected at the M parent on the same eligible positions, while the native removal background retains every other visible continuation term. Native Full reuse verifies the exact parent, all 1,024 row/mask/denominator steps, optimizer clock and validated numerical kernel, not merely ordinary-CE algebra. The additional controls comprise two full-pool, two offset and four native-removal trajectories.

Saved state and time-course diagnostics. A comparison of all 8,192 saved S/M risk measurements matches exact article id, eligible position and prefix hash. Spearman uses average ranks for ties. The unweighted mean native h is .065915/.066291 for S and .003275/.002841 for M (orders 2/3); these means are not the occurrence- and D_B-weighted scalar sums in Table 3. H-set intersections are 787/783, with Jaccard .922626/.913652 and S/M continuation-loss rank correlations .994915/.994979. This describes state-associated profiles, not a parameter mechanism. The time-course analysis covers four arms per order: full-pool random, offset, and the two native removals. Their 1,024-update trajectories use logical batches of 16 and microbatches of two. Each update contributes a pre-update minibatch measurement, giving 8,192 update observations. Scalar mass is $\sum_j \lambda w_j \sigma(a_j + d_j)/D_B$, distinct from a gradient norm or accumulated Adam update. At anchors 0/32/128/256/1,024, eight fixed training rows (four stop and four non-stop events) are probed at boundary and retained positions, giving 640 observed-prefix measurements. Four fixed training prompts per order use parent-specific 128-token greedy sequences generated once with EOS suppressed. Probing their fixed prefixes across arms and anchors gives 160 first-stop summaries,

Table 6: First-step total-update diagnostics. Median [first, third quartile] over 32 fixed training batches within each selective parent. Batches are not independent training runs; u_A denotes the complete fresh-AdamW displacement.

Quantity	Order 2	Order 3
$\|u_H\|_2$	.401035 [.400791,.401134]	.400936 [.400731,.401162]
$\|u_{R_{\text{pool}}}\|_2$	.401035 [.400786,.401153]	.400936 [.400732,.401150]
$\|u_{R_{\text{offset}}}\|_2$	.401034 [.400777,.401153]	.400930 [.400732,.401154]
$\cos(u_H, u_{R_{\text{offset}}})$	.972133 [.944306,1.000000]	.954196 [.928865,1.000000]
$\|u_H - u_{R_{\text{offset}}}\|_2$	.094601 [5.80×10^{-6},.133837]	.121390 [2.28×10^{-5},.151364]

not free-running outputs. The first-stop index is zero-based, with no-stop recorded if all 128 prefixes continue. S-background arms share S sequences and M-background arms share M sequences; common prompt IDs do not imply common token sequences. These repeated diagnostics are not independent samples. The online offset sums visit changing minibatches and changing parameters; no historical Adam state or missing H trajectory is reconstructed. Native probe a/h excludes the surrogate offset. Offset online scalar sums are .440465/.429324 versus initial matched sums .658559/.652699. All 1,024 updates of each of the eight additional training trajectories exceed preclip norm 1 for the total gradient, not continuation loss alone. Four fixed training-prefix sequences per order retain their initial first-stop index at 0/32/128/256/1,024 for offset. Eight unchanged prefix-sequence stops do not imply unchanged free-running generation.

Initial local-update diagnostic. A separate diagnostic uses 32 training batches at each selective parent, with indices $\lfloor k\,1023/31 \rfloor$ for $k = 0, \ldots, 31$ fixed before measurement. Selection uses training inputs and parent metadata, without effect-size selection. Each batch restores parent weights and RNG, computes Q+STOP and the H/R_{pool}/R_{offset} continuation-loss gradients, and compares the fresh-AdamW first displacements before restoring the parent. Across 64 batches, the diagnostic evaluates 192 candidate first steps without advancing the parent weights. No new endpoints or generated outputs result. CPU recomputation verifies the assembled scalar and coordinate calculations, not the full neural backward pass. All partially processed batches from two calculations stopped by projected-runtime limits are excluded; the reported results cover the complete fixed 64-batch roster. These local measurements do not identify the cause of the 1,024-update generation effects or provide trajectory-wide dose matching.

Full-parameter FP64 reductions cover all 1,711,376,384 unique coordinates; separate 32-coordinate checks are not extrapolated into full norms. Table 6 reports within-parent medians and quartiles. The Q+STOP gradient was computed, but its optimizer-only displacement u_0 was not stored, so isolated continuation increments $u_A - u_0$ cannot be reconstructed. Full-pool random is not a substitute for u_0. Total displacements include common content/STOP contributions and decoupled decay; high total-update cosine does not establish aligned continuation-only increments.

Fresh-source native analogue. XSum uses disjoint sets of 32,768 warm-start and 16,384 continuation rows, two fresh native 2,048-update parents and three 1,024-update continuations per parent. The native-complete eligible universe contains 452,513 observed preterminal positions; H selects the prespecified upper tenth within its own parent, and full-pool random matches the fixed strata. All content and true STOP remain; Full equals native CE under the validated factorization. No synthetic censoring creates an effect. Source-only prompt eligibility has maximum 1,536 tokens and training mean 510.458; selecting 256 evaluation articles examines 272 source records and excludes 16 by length, without outcome-based refill. Full/H/random ROUGE-L means are .310014/.308630/.309146 and lengths 24.846/24.637/24.711. The shared 256 articles, not 512 independent rows, are resampled across both orders. Dataset provenance and role hashes do not establish global unseen status, pretraining exclusion or redistribution rights.

B Complete retention contrasts

The table below reports all corners needed for the H and R finite differences on the fixed 256 articles. Values average the same two training orders, with no arm-specific article filtering.

Table 7: Absolute flat ROUGE-L, fixed 256-article panel.

Endpoint	Mean	Order 2	Order 3
Content + STOP	.217742	.214743	.220741
Only-H	.286564	.283156	.289972
Only-R	.217178	.214676	.219681
Drop-H	.217230	.214143	.220317
Drop-R	.288774	.283882	.293665
Full support	.288241	.285384	.291098

Table 8: Paired 256-article contrasts. All intervals are 95%; J and the Q-hard contrast are secondary. The primary is only-H − only-R.

Contrast	Difference	Interval
Δ_H^{add}	+.068822	[.058699,.079143]
Δ_H^{remove}	+.071011	[.061072,.081404]
J_H	+.002190	[−.000033,.004526]
J_R	+.000031	[−.002214,.002242]
Only-H − only-R	+.069385	[.059588,.079814]
Only-H − Q-hard	+.067867	[.057958,.077811]
Only-H − full	−.001678	[−.003808,.000285]

The training-only profile and fixed-offset surrogate appear in Table 3 and Section 5. The original retention intervals above are unchanged; later controls do not establish parameter-gradient equivalence.

Table 9: Additional control contrasts in flat ROUGE-L. All intervals are descriptive 95% article-paired intervals; orders are fixed, not independent pretraining replications. The XSum panel is a native-complete analogue.

Panel	Contrast	Mean [95% interval]	Order 2	Order 3
256	H − R_{pool}	.058827 [.049060,.068722]	.062047	.055607
	$R_{\text{pool}} - R$	.010558 [.006287,.015124]	.006433	.014684
	$R_{\text{pool}} - \varnothing$	.009995 [.005660,.014414]	.006365	.013624
	$R_{\text{offset}} - R$	−.000236 [−.002343,.001781]	−.000266	−.000206
	Native dropR − dropH	−.000502 [−.002674,.001663]	.000860	−.001864
1,536	H − R_{pool}	.057129 [.053339,.060894]	.057232	.057026
	$R_{\text{pool}} - R$	.010348 [.008758,.011987]	.006403	.014293
	$R_{\text{offset}} - R$	.000542 [−.000377,.001530]	.001334	−.000250
	Native dropR − dropH	.000322 [−.000561,.001234]	.000396	.000247
XSum 256	H − R_{pool}	−.000516 [−.002845,.001744]	.000680	−.001713
	H − Full	−.001384 [−.003585,.000605]	−.002270	−.000498

Fixed-support parent crossover. The later crossover transfers the original S-selected H_S and $R_{\text{pool},S}$ masks to each native-objective parent M without reranking or redrawing. Both parents receive the same selective continuation background: Q, complete STOP and the indicated eligible continuation support. Eligible full contains the original 8,192 eligible continuation positions; full support under the native objective includes every visible continuation term and remains a separate comparator. Six M continuations use the matched 1,024-update tape and fresh optimizer; six validated S endpoints are reused. Results for the S parent and full-support native objective on these same 256 articles were already available before the new M continuations. Masks, continuations and contrasts were fixed before the M continuations. The local-update diagnostic instead uses the original training tapes. Prior-result exposure before these analyses was not fully recorded. Both analyses remain descriptive rather than outcome-blind new confirmations. M0 denotes the M parent without continuation training. The fixed 256 articles and both orders remain linked in 4,096 bootstrap draws (seed 2026092543). This exposed, hypothesis-driven analysis has descriptive 95% intervals and no independent pretraining replication or 1,536-article extension. The primary interaction is the M − S

difference in $H_S - R_{\text{pool},S}$. The reused S effect has a slightly different descriptive interval under this bootstrap seed; the original retention intervals remain unchanged.

Table 10: Complete crossover flat ROUGE-L contrasts on 256 shared articles. Intervals are descriptive 95%; H and R_{pool} retain S-defined membership under both parents. Relative selection advantage does not establish improvement over the uncontinued M0 parent.

Contrast	Mean [95% interval]	Order 2	Order 3
Parent interaction (M − S)	−.036458 [−.045404,−.027444]	−.034257	−.038658
S: H − R_{pool}	.058827 [.048809,.068925]	.062047	.055607
M: H − R_{pool}	.022369 [.015979,.029293]	.027790	.016948
M: H − Eligible full	−.001334 [−.004212,.001486]	−.001746	−.000921
M: R_{pool} − Eligible full	−.023703 [−.031101,−.016746]	−.029537	−.017870
M: Eligible full − native Full	.005640 [−.001767,.012859]	.004999	.006281
M: H − M0	.006803 [−.002349,.015981]	.008345	.005260
M: R_{pool} − M0	−.015567 [−.027142,−.003879]	−.019446	−.011688
M: Eligible full − M0	.008136 [−.001716,.017832]	.010091	.006182
Native Full − M0	.002496 [−.005301,.010022]	.005092	−.000100

Complete paired measurements retain length, EOS and cap incidence, overlap and source support, including adverse outcomes. The M random continuation falls below M0, while H's improvement over M0 is inconclusive. The crossover therefore strengthens the regime boundary without converting relative selection performance into an absolute benefit or a parameter mechanism.

C Generation crosses and the larger confirmation

The larger-panel conditions are reported in Table 4. The following details specify how native, crossed, and exact-token forced outputs remain comparable without treating their different stopping rules as natural interventions on training.

Decoder semantics. Each decision donor and content donor sees the same evolving prefix at each step. The raw-greedy rule preserves the configured EOS/EOT set and native tie behavior. On a continue decision, the content donor contributes the highest-probability non-stop token. A diagonal uses the same model for both roles and must reproduce the corresponding native output. A forced-length cross uses the paired native donor's generated content-token count, excluding terminal markers, never the reference summary length. Ignoring an earlier predicted stop for this diagnostic changes the trajectory and is not a deployable factuality correction.

Termination and cost. All native selective conditions in this panel terminate at EOS without a cap. Native M has cap fraction .007487 and EOS fraction .992513. Forced-length rows are marked as forced rather than naturally EOS-terminated. The two donor forwards are charged as two model evaluations. Retention does not reduce the input tape, Q computation, or update count; the support mask is a scientific intervention, not an implemented speedup.

Conditional uncertainty. For each bootstrap draw, article indices are sampled with replacement and the corresponding outputs from every arm and both orders remain linked. There are 4,096 draws. The 256-panel gate/content intervals retain their 97.5% level; the larger panel's secondary intervals are descriptive 95% intervals, without simultaneous adjustment across all metrics. The larger only-H − only-R contrast checks the same fixed difference on more articles without changing the earlier primary, creating new trained endpoints, or establishing population-level training uncertainty.

C.1 Complete paired decoder and boundary comparisons

These tables report the paired decoder and boundary comparisons. Values come from the same fixed paired measurements and are rounded only for display. The intervals condition on the two fixed orders and are not uncertainty over independent pretrained models.

Table 11: Complete paired effects in flat ROUGE-L, 1,536 articles. Intervals are descriptive 95% article intervals, without adjustment across secondary metrics. Orders 2 and 3 remain separate.

Contrast	Mean	95% interval	Order 2	Order 3
Removal total	$+0.068576$	$[+0.064519, +0.072920]$	$+0.065451$	$+0.071702$
Gate transfer	$+0.067504$	$[+0.063377, +0.071801]$	$+0.064372$	$+0.070637$
Content at R gate	$+0.001072$	$[+0.000068, +0.002115]$	$+0.001079$	$+0.001065$
Content at long length	$+0.001120$	$[+0.000149, +0.002127]$	$+0.000711$	$+0.001530$
Content at short length	$+0.000270$	$[-0.000394, +0.000954]$	$+0.000133$	$+0.000407$
Full − native M	-0.006326	$[-0.009964, -0.002746]$	-0.012369	-0.000284

Table 12: Complete paired effects in automatic source-support proxy, 1,536 articles. Intervals are descriptive 95% article intervals, without adjustment across secondary metrics. Orders 2 and 3 remain separate.

Contrast	Mean	95% interval	Order 2	Order 3
Removal total	-0.024824	$[-0.032093, -0.017184]$	-0.021022	-0.028625
Gate transfer	-0.022973	$[-0.030122, -0.015851]$	-0.019732	-0.026214
Content at R gate	-0.001851	$[-0.004286, +0.000562]$	-0.001289	-0.002412
Content at long length	$+0.005403$	$[+0.002646, +0.008005]$	$+0.006443$	$+0.004363$
Content at short length	-0.005588	$[-0.008579, -0.002618]$	-0.006657	-0.004519
Full − native M	$+0.011562$	$[+0.006544, +0.016581]$	$+0.012320$	$+0.010803$

C.2 Exhaustive training-calibrated stopping-bias control

This inference-time control differs from initial training-pressure matching. The drop-H content model, weights and all non-STOP logits remain fixed; a single FP32 bias is added to EOS token 2. Each fixed training order uses 64 training articles and 512 non-STOP prefixes per article. The donor's first stopping length is the calibration target. Evaluation references and outputs do not select the bias.

For each training prefix, the search identifies the first FP32 bias that makes EOS win under the native argmax tie rule. Bias -64 continues and $+64$ stops at every prefix. Evaluating every threshold, its preceding representable value and zero covers all first-stop behaviors of this constant-bias family on the fixed training sequences. Selection minimizes total article-wise absolute length error to donor stopping; ties favor smaller absolute bias and then the smaller signed value. Exhaustiveness concerns this training length objective, not evaluation ROUGE or every scalar stopping controller.

Table 13: Training-only bias selection. The earlier ten-candidate quantile grid selected zero; exhaustive search supersedes that weak scalar comparator. MAE is in tokens.

Order	Coarse-grid MAE at zero	Selected bias	New MAE
2	23.59375	-7.4654211998	14.34375
3	31.37500	-8.1736698151	15.640625

The selected biases generate 512 new outputs on the same 256 exposed articles and two fixed training orders. Decoding uses FP32 scaled-dot-product attention, TF32 disabled, greedy selection and a 512-token cap. Zero-bias decoding reproduces native outputs before applying the intervention. Comparator outputs reuse the existing native and donor decoders.

The global correction recovers approximately 67.94% of the native-to-donor overlap gap, a descriptive ratio. Donor stopping remains better in each order. Near recovery of mean length does not imply article-wise length matching: mean absolute length error remains 18.36 tokens and exact output agreement is 38.09%. These results therefore neither isolate content quality at matched article lengths nor establish factuality or an advantage over ordinary native training. The comparison concerns a length-calibrated global constant, not richer input-dependent controllers or other calibration objectives. Intervals condition on the endpoints, selected biases and exposed articles, excluding independent training variation. New calibration and generation use .816501 worker GPU-

Table 14: Global-bias comparison on the exposed 256 articles. ROUGE-L is multiplied by 100; length excludes EOS. Length error and exact token-output agreement use the donor-stopping decoder as reference. All arms have EOS rate 1 and cap rate 0.

Decoder	ROUGE-L	Length	Length MAE	Agreement (%)
Native drop-H	21.723013	21.35742	32.93750	5.46875
Training-selected bias	26.553867	52.75195	18.36328	38.08594
Donor stop, drop-H content	28.833291	54.29492	0	100
Native drop-R	28.877367	54.04688	1.70508	81.05469

Table 15: Global-bias contrasts in ROUGE-L percentage points. The separate two-contrast family uses 97.5% intervals from 4,096 paired article resamples, averaging the two fixed order effects within each article; random seed 2026092617.

Contrast	Difference	Adjusted interval
Bias − native drop-H	$+4.830854$	$[+3.793023, +5.881689]$
Bias − donor stopping	-2.279423	$[-3.123708, -1.459471]$

hours (.821382 including allocated GPU elapsed time); earlier training trajectories and comparator generation are reused, not cost-free.

D Measurement sensitivity and reproducibility

Source-support uses the fixed official AlignScore-base inference path with its prescribed sentence/chunk scoring. Highlight support queries literal reference-highlight lines against the generated response. The two directions have different denominators and are not interchangeable measures of fact coverage. BERTScore and ROUGE may provide additional similarity measures, but the central claims do not depend on treating either as factual ground truth. No new human annotation is asserted.

Table 16: Earlier long-continuation source-support confirmation. Both order pairs use the same 128 articles. The primaries are inconclusive; positive secondary interactions are not independent confirmations of content improvement.

Order pair	Estimand	Difference	Interval
0/1	Content primary, long length	+.010868	[−.004528,.025731]
2/3	Content primary, long length	+.004954	[−.011998,.021046]
0/1	Length × content interaction	+.023341	[.003978,.042863]
2/3	Length × content interaction	+.030384	[.012523,.049369]

The long-length primary compares the long repair endpoint with the content/STOP endpoint forced to the repair endpoint's native token count; both received 3,072 extra updates. The original orders 0/1 and later orders 2/3 use the same 128 articles. Each analysis averages the two order effects within article and draws 4,096 paired resamples. Endpoint, metric and stopping interventions differ from the planned-secondary content-at-drop-R-gate contrasts on 256 or 1,536 articles.

The primary intervals above are 97.5%; secondary interaction intervals are 95%. The interaction is the long-length content contrast minus the short-length content contrast. This interaction can be positive even if a positive content-improvement primary is not established.

The paired-complete short-length comparison retains 114/128 articles and excludes 14; the long-length comparison retains 125 and excludes 3. The projected differences are +.000238 [−.000133,.000864] and −.000641 [−.010624,.008668]. The paired-complete native removal comparison retains 121 articles and gives −.048120 [−.074569,−.020309]. These projections use saved sentence scores, with no sentence completion, paraphrasing, or new semantic model inference. Their conditional subsets and changed lengths prevent promotion into replacements for the exact-token primary.

Table 17: Sentence-completion sensitivity on the earlier 128-article removal panel. Counts are order 2 / order 3. Zero-complete rows are not assigned an invented support score. Forced trimming is not an equal-token comparison.

Condition	Final fragments	Zero complete sentences
Drop-H	7 / 5	7 / 5
Drop-R	7 / 5	3 / 0
Drop-H at drop-R length	22 / 21	3 / 0
Drop-R at drop-H length	12 / 10	10 / 9

D.1 Evidence-to-claim map and reproducibility

The support claim uses the 256-article complement and Q-hard comparisons; the stopping explanation uses paired decoder crosses, the larger-panel comparison and the exhaustive global-bias control. Initial pressure matching is assessed from training-only profiles. Measurement limits follow from the failed evaluator tests and sentence-completion sensitivity. Reported comparisons retain fixed trained weights, masks and article panels, with native-output identity checked before decoder contrasts. Adverse outcomes remain reported; invalid measurements and unfinished analyses do not enter effect estimates.

The following limitations remain explicit: no specified noninferiority margin; no trajectory-wide dose or parameter-direction matched training; no validated atomic fact inventory; no independent pretraining replication; and no proof of global historical or pretraining non-exposure. The uncertainty estimates remain conditional on the trained models and do not cover between-training variation.