ANSWER FIRST, REASON LATER: WHEN COMMITMENT ORDER COSTS ACCURACY IN DIFFUSION LANGUAGE MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Masked diffusion language models write a response by filling a fixed-length canvas of masked positions in any order, so the final answer can be fixed before the reasoning printed above it exists. It is not known when this freedom of order costs accuracy, or whether the early answer is the reason. We record when every position of a response commits and compare unrestricted confidence-based decoding with frontier gating, which lets only positions near the left-most unfinished one commit, on GSM8K and MATH-500 with LLaDA and Dream models. In LLaDA models the cost lies entirely in *end-anchored* trajectories, in which the end of the response, including the answer, commits before its middle: gating adds $+32.0$ points of accuracy on these GSM8K responses and $+10.7$ on MATH-500, and nothing elsewhere, where unrestricted decoding already proceeds from left to right. Every answer-first response is end-anchored. End-anchoring can be predicted before decoding from the share of the canvas that the first forward pass assigns to end-of-sequence tokens (AUC 0.83–0.94), and it becomes more frequent with step-by-step prompting and harder problems and less frequent with a longer canvas, which accounts for the prompt and canvas dependence of the gain. Two pre-specified interventions test this account. Holding only the last 32 positions of the response back until step 448 of 512 removes end-anchoring and recovers the whole gain of gating on end-anchored GSM8K questions ($+31.4$ against $+32.0$ points), whereas holding an equally long span in the middle changes nothing; conversely, widening the gate until the end of the canvas becomes available brings end-anchoring back and removes the gain (-8.1 points). Delaying only the answer leaves 90% of end-anchored responses end-anchored and recovers little. Evaluations of diffusion language models should report canvas length, termination rule, and commitment order alongside accuracy.

1 INTRODUCTION

Reasoning prompts commonly request intermediate steps followed by a final answer (Wei et al., 2022). In autoregressive generation, this presentation order is also the order in which tokens are produced, so the reasoning exists when the answer is written. Masked diffusion language models separate the two orders: they fill a fixed-length *canvas* of masked positions in any order, and a position *commits* when its token becomes final, so tokens near the end can be fixed before those near the beginning (Austin et al., 2021; Nie et al., 2025). A prompt can therefore specify where reasoning should appear without specifying whether that reasoning exists when the answer is committed.

Under the decoding procedures studied here, commitment is irreversible, and a complete answer can commit while most of the reasoning displayed before it is still masked (Figure 1). Tokens committed afterward cannot revise that answer, even though the finished response presents them as the steps that led to it. The displayed reasoning therefore does not show how much of it existed when the answer was fixed. Commitment traces make this order observable; they do not reveal the timing of internal computation or whether the explanation is semantically faithful, which requires interventions on the reasoning itself (Turpin et al., 2023; Lanham et al., 2023).

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.

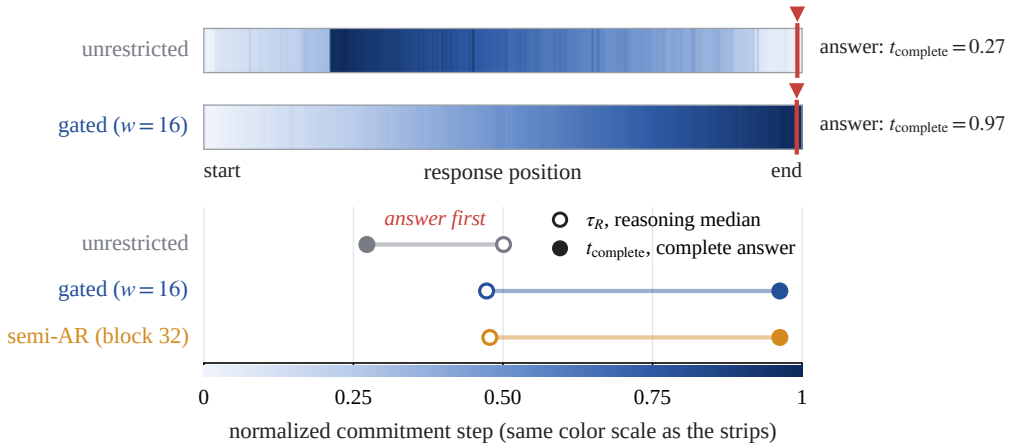

Figure 1: **Displayed last, committed first.** Top: a representative MATH response under unrestricted and gated decoding. Color shows the normalized step at which each response position commits; the red marker is the answer span. Unrestricted decoding fixes the complete answer while most of the canvas is still masked and then fills the reasoning in behind it (an *end-anchored* trajectory); gating commits the answer last. Bottom: medians of the complete-answer step and of the median reasoning step over MATH-500 outputs with a complete mapped answer span (LLaDA-8B, L=512; 240–246 responses per policy).

Prior work establishes that diffusion models can decouple requested output order from generation order. Yu et al. (2026a) examine robustness to answer–reasoning presentation order, Ni et al. (2026) show that confidence-based generation can bypass uncertain reasoning choices and reduce solution coverage, and decoders use positional restrictions and deferred commitment to improve quality (Shu et al., 2026). That restricting commitment order can help reasoning is therefore known; what an aggregate improvement leaves open is which errors a restriction repairs, how the benefit depends on generation conditions, and whether the early answer is what causes the loss.

We address these questions by joining commitment traces to the paired outcomes of two decoders under a fixed requested answer format: unrestricted confidence-based decoding and *frontier gating*, which allows only the 16 positions after the left-most unfinished position to commit. Gating serves as a diagnostic, not as a proposed method; the blockwise decoder released with LLaDA performs comparably. In LLaDA models, the gain from gating turns out to lie entirely in *end-anchored* trajectories, in which unrestricted decoding commits the end of the response, including the answer, before its middle; all other responses are already decoded from left to right, so the restriction changes nothing. End-anchoring includes every answer-first trajectory, can be predicted from the first forward pass, and varies with the prompt, the difficulty of the problem, and the canvas length in the way the gain does. Two pre-specified interventions bracket the effect: holding back only the tail of the response removes end-anchoring and recovers the whole gain of gating, and widening the gate until the end of the canvas becomes available brings end-anchoring and the accuracy loss back together, whereas delaying commitment of the answer alone leaves the anchor in place.

Accuracy under unrestricted decoding thus depends on the prompt, the canvas length, and the termination rule through the share of end-anchored trajectories, so these settings and the commitment order should be reported with accuracy; for users, holding back the end of the response or restricting the order removes the cost and a larger canvas reduces it, whereas postponing the answer does not.

Our contributions are:

- **Where the cost lies.** On GSM8K, 18.1% of LLaDA-8B trajectories are end-anchored; gating raises their accuracy from 32.0% to 63.9% and changes it by -0.2 pp elsewhere. On MATH-500 with a 512-position canvas the gains are $+10.7$ pp and $+1.2$ pp. Outside end-anchored trajectories unrestricted decoding is already left to right, and the split holds within difficulty strata and for LLaDA-1.5.

- **When it occurs.** Every answer-first trajectory is end-anchored, and end-anchored responses fill the canvas. The share of the canvas assigned to EOS by the first forward pass predicts end-anchoring (AUC 0.83–0.94; 87.5% accuracy on held-out GSM8K questions) and separates the gain. Step-by-step prompting raises the end-anchored share from 18.1% to 26.2% without changing the gain within it, which accounts for the $+3.3$ pp prompt–policy interaction, and a longer canvas lowers the share and the gain together; Dream-7B rarely end-anchors and fails differently.
- **What removes it, and what brings it back.** On the 194 GSM8K questions that are end-anchored under unrestricted decoding, holding the last 32 positions of the response ineligible until step 448 of 512 removes end-anchoring and raises accuracy by $+31.4$ pp, as much as gating ($+32.0$ pp), while holding an equally long span in the middle gains $+0.5$ pp. Widening the frontier to the canvas length restores end-anchoring (0% to 81% of trajectories) and lowers accuracy by -11.0 points within problems. Delaying the answer position alone gains little ($+0.8$ pp on open-ended GSM8K, where gating gains $+5.6$ pp), and 90% of end-anchored responses stay end-anchored.

2 METHODS

2.1 MODELS, TASKS, AND DATA SPLITS

We study three instruction-tuned diffusion language models: LLaDA-8B-Instruct, pretrained from scratch with masked diffusion and then supervised fine-tuned (Nie et al., 2025); LLaDA-1.5, which adds VRPO preference optimization (Zhu et al., 2025a); and Dream-7B, adapted from an autoregressive model to discrete diffusion with context-dependent noise scheduling (Ye et al., 2025). All three are decoded with the same low-confidence remasking loop (one forward pass per step, a fixed number of commitments per step, and the same frontier gate); Dream-7B additionally uses its next-token logit alignment, and its native sampler is not used. Cross-model contrasts are descriptive because these models differ in several training choices.

The main tasks are GSM8K (Cobbe et al., 2021) and MATH-500 (Hendrycks et al., 2021; Lightman et al., 2024). Unless noted otherwise, the output canvas has $L{=}512$ positions, decoding uses 512 denoising steps at temperature zero, and the released low-confidence remasking procedure is used with EOS/EOT confidence suppression enabled, applied to each model's own terminator tokens. All primary results condition on this termination rule (Appendix G).

The first 250 GSM8K test questions establish the prompt comparison and trajectory analysis, excluding three inspected during interface development ($n{=}247$). Applying the same analysis to the remaining 1,069 questions supplies the main estimate. The model-by-canvas MATH study similarly uses the first 250 MATH-500 problems minus these three indices, leaving 247. A separate $L{=}512$ evaluation uses the remaining 250 problems (indices 250–499).

2.2 COMMITMENT POLICIES AND SCHEDULES

Unrestricted decoding allows every masked position to commit. Frontier gating permits only positions $i < f(t) + w$, where $f(t)$ is the left-most unresolved position at step t; the default width is $w{=}16$. Semi-autoregressive decoding completes blocks of 32 positions from left to right. These *eligible sets* vary while the model, prompt, temperature, scoring rule, and total commitment budget stay fixed. Each policy produces a different intermediate context, so later token scores and reasoning trajectories can differ.

We use frontier gating as a fixed-width diagnostic, not as a proposed decoder: semi-autoregressive decoding, the blockwise procedure released with LLaDA, performs comparably wherever we evaluate both (Tables 2 and 7), so accuracy differences measure the cost of unrestricted confidence ordering.

In the canvas-length study, the default number of denoising steps equals L. At $L{=}1024$, the 1,024-step schedule allocates one commitment per step; the 512-step schedule allocates two and matches the function-evaluation count at $L{=}512$. The realized median counts are one and two, respectively (Appendix B).

2.3 ANSWER MAPPING AND TRAJECTORY MEASUREMENTS

For output position i, let $\tau(i)$ be its normalized first irreversible commitment step. GSM8K answers are located with the fixed declared-answer extraction cascade used by the scorer. MATH expressions are mapped with a brace-balanced parser that selects the last syntactically complete `\boxed{...}` expression, including nested braces, or an unambiguous declared complete expression. A final-number fallback is not treated as a complete MATH answer span. Outputs without a primary complete mapping retain their correctness label but receive no answer-timing or backfill value. MATH trajectory summaries additionally require at least 20 generated tokens before the answer and defined answer and reasoning times. Paired comparisons use the intersection of eligible items; mapping coverage appears in Appendix F.

Two annotators independently reviewed 183 stratified outputs, oversampling fallback and boundary cases. Joining their spans to the commitment-step arrays allows direct remeasurement of t_{complete}, answer-first status, and backfill (Appendix F). These spans also informed the parser correction, so agreement is measured on a development sample, not held-out data.

For a mapped answer-token set A, the primary answer time is

$$t_{\text{complete}} = \max_{i \in A} \tau(i), \tag{1}$$

the commitment time of the last answer token to become irreversible, equivalently the earliest step at which the entire extracted answer is fixed. We retain the median answer-token time $t_{\text{med}} = \text{median}_{i \in A}\tau(i)$ as a robustness statistic. For reasoning tokens R printed before the answer, the comparator $\tau_R = \text{median}_{i \in R}\tau(i)$ locates the center of commitment while limiting the influence of isolated early or late tokens; Appendix F discusses earliest- and latest-reasoning alternatives. The order gap is

$$g = t_{\text{complete}} - \tau_R, \tag{2}$$

so $g < 0$ denotes answer-first commitment: the complete answer is fixed before the median reasoning commitment, not necessarily before every reasoning token. Correction analyses define answer-first status from the unrestricted trajectory before examining the gated result.

We separately record an *early-answer-start* diagnostic when fewer than 20 generated tokens precede the mapped answer. This is a positional category, not a claim that the response contains no reasoning. GSM8K correction groups exclude these early starts and outputs without a defined reasoning region. A response, the content before the first terminator token on the canvas, is *canvas-filling* when its length is within four tokens of L, and *end-anchored* when the median commitment step of its last 16 tokens precedes that of the central half of the response; Appendix C varies both choices. The *step-0 EOS share* is the fraction of canvas positions whose most likely token in the first forward pass is one of the model's terminator tokens (end of sequence or end of turn); it is computed before any commitment, from the raw predictions to which EOS/EOT confidence suppression has not yet been applied. For MATH reasoning positions R, *backfill* is the fraction committed after the complete answer, $|\{i \in R : \tau(i) > t_{\text{complete}}\}|/|R|$; it describes temporal allocation, not logical validity.

Prompt–policy comparison. GSM8K uses a direct-answer instruction, a format-only instruction that permits reasoning, and an explicit instruction that adds "Let's think step by step." The prompt–policy interaction is the gain from gating under the explicit instruction minus the gain under the format-only instruction.

2.4 CONTROLLED ANSWER DELAY

The controlled task presents four numerical options per GSM8K question. Distinct distractors are training-set answers of the gold value's order of magnitude; a fixed permutation assigns single-token A–D labels. The canvas reserves 280 masked reasoning positions and one masked answer position between immutable delimiters and a termination tail. This scaffold specifies answer width and location before generation, defining a different task from open-ended GSM8K.

A free reference pass records each position's commitment time. *Answer delay* keeps the answer position masked and ineligible for commitment until release step r_i while every other position commits as usual; the model still predicts the position but cannot fix it. The *matched reasoning delay* instead holds back a reasoning position that commits before r_i and is closest to the answer's reference

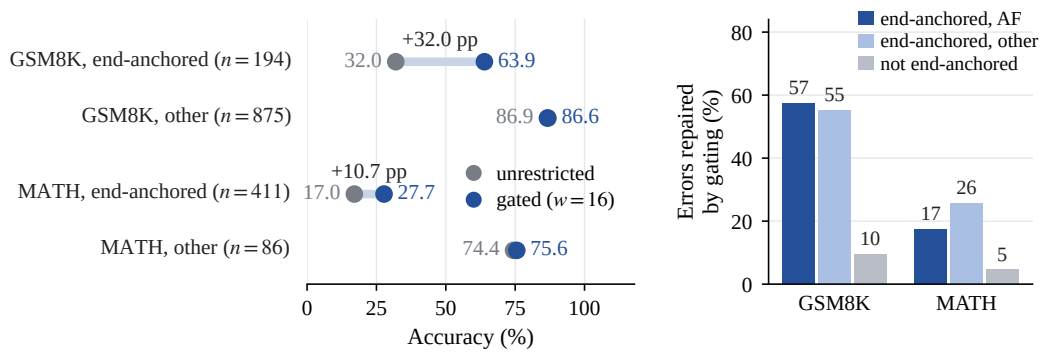

Figure 2: **Gating repairs errors only in end-anchored trajectories.** Left: accuracy on the same problems under unrestricted and gated decoding, split by whether the unrestricted trajectory is end-anchored (LLaDA-8B, L=512, format-only instruction; GSM8K: remaining 1,069 questions; MATH-500: pooled 497 problems). Right: share of unrestricted errors that gating repairs. Within end-anchored trajectories, answer-first (AF) errors are repaired no more often than the others. Group comparisons are observational.

commitment time. Selection uses neither correctness nor intervention outcomes. Both conditions manipulate one position, preserve the per-step commitment budget, and supply no answer content.

The main comparison releases at each item's reference reasoning-median step, nominally 140 of 281; earlier and later releases test schedule sensitivity. A free-form replay instead targets positions obtained from a completed reference answer, with matched reasoning targets; it tests a reference-defined positional restriction rather than a slot known in advance. Appendix E gives the construction, the manipulation checks, and further controls (a pre-specified random-position delay and an influence-matched delay). Within every paired comparison of decoding policies, both arms were generated by the same sampler build on a single GPU class; the two exceptions among the delay experiments are documented in Appendix B.

Scoring and uncertainty. GSM8K uses normalized numeric exact match. MATH combines normalized string equality with symbolic verification, applied identically across policies and independently of answer-span mapping. The controlled task scores the final option label. Paired binary contrasts use percentile bootstrap intervals and exact McNemar tests. Joint interactions resample each question's outcomes together. Main-text intervals are 95% unless labeled 90%. Resampling counts, seeds, and exact tests appear in Appendix B.

3 Results

3.1 The gain lies in end-anchored trajectories

We first ask which errors frontier gating repairs. The GSM8K comparison crosses format-only and step-by-step instructions with unrestricted and gated decoding on the remaining 1,069 questions, holding the final-answer format fixed; MATH-500 results use the conditions of Section 3.2.

End-anchored trajectories. Under the format-only instruction, 18.1% of unrestricted GSM8K responses are end-anchored. Gating raises accuracy on these questions from 32.0% to 63.9% ($+32.0$ pp, 95% CI [+23.6, +40.3]) and leaves the other 875 unchanged (-0.2 pp, 95% CI [-1.3, +0.9]; Figure 2, Table 1). On MATH-500 with a 512-position canvas the gains are $+10.7$ pp and $+1.2$ pp, and LLaDA-1.5 shows the same split on both benchmarks. The split is not a difficulty effect: end-anchoring is more frequent on harder problems, but within every MATH difficulty level, and within every third of the GSM8K questions ordered by the number of steps in the reference solution, the gain lies in end-anchored trajectories (Figure 8). Outside them the restriction is not binding: unrestricted decoding already commits those responses from left to right (median position–order correlation 0.995; Appendix A).

Table 1: **Where gating helps.** Trajectories are split by whether the unrestricted response is end-anchored. The overall gain from gating is approximately the end-anchored share times the gain within that group, because gating changes the outcome of few other problems (last column: wrong-to-correct / correct-to-wrong transitions). GSM8K rows use the remaining 1,069 questions (†the first 247); *pooled 497 MATH problems; other MATH rows use 247. Figure 7 shows every condition.

		Overall	End-anchored trajectories		Other trajectories
Task, model	Condition	gain (pp)	Share	Gain within (pp, 95% CI)	Outcomes changed
GSM8K, LLaDA-8B	format-only	$+5.6$	18%	$+32.0$ [+23.6, +40.3]	24 of 875
	step-by-step	$+8.9$	26%	$+33.6$ [+26.8, +40.2]	25 of 789
GSM8K, LLaDA-1.5	format-only†	$+13.4$	28%	$+45.7$ [+31.9, +59.1]	5 of 177
MATH, LLaDA-8B	L=256	$+9.3$	96%	$+9.2$ [+3.8, +14.9]	1 of 9
	L=512*	$+9.1$	83%	$+10.7$ [+6.6, +14.9]	1 of 86
	L=768	$+1.2$	57%	$+2.8$ [-3.6, +9.4]	3 of 105
	L=1024	$+0.4$	44%	$+0.9$ [-7.4, +9.3]	6 of 138
MATH, LLaDA-1.5	L=256	$+13.0$	99%	$+13.1$ [+7.0, +19.1]	0 of 3
	L=512	$+6.9$	83%	$+7.3$ [+0.5, +14.2]	2 of 42
	L=768	$+6.9$	60%	$+13.6$ [+6.7, +20.7]	5 of 100
	L=1024	$+4.9$	49%	$+9.8$ [+2.3, +17.6]	6 of 125

Is it answer-first commitment, or a full canvas? Two simpler descriptions of these trajectories are available, and neither explains the gain better. *Answer-first commitment:* all 139 answer-first outputs on GSM8K are end-anchored, which is why they are 13.6% of eligible responses but 54 of the 86 repaired errors (Table 11); yet the 55 end-anchored outputs that are not answer-first gain as much ($+29.1$ pp against $+33.1$ pp). *Canvas filling:* every end-anchored response fills the canvas, because content fixed at its end leaves no room for an earlier stop; yet canvas-filling responses that are not end-anchored show no clear gain ($+3.4$ pp, 95% CI [-17.2, +24.1]; Table 3), and gated responses fill the canvas on half of the MATH problems without any loss.

What is fixed first. In end-anchored responses the closing tokens around the answer commit first (median at 0.09 of the schedule), the answer follows (0.35), and the middle of the reasoning comes last (0.71). The answer is thus committed before the solution exists, and while the model's own prediction of it is still changing. One step before the first answer token commits, we compare the model's most likely token at each answer position with the token that ends up there. They match at 97% of positions in other trajectories, where the reasoning above already determines the answer, but at only 71% in end-anchored ones, and at 94% for the same questions under gating (Table 4).

3.2 End-anchoring follows the canvas, the prompt, and the first forward pass

We next ask when end-anchoring occurs. On MATH-500 we vary the canvas: a short-canvas comparison on 497 non-development problems for LLaDA-8B, a 247-problem, three-model comparison across canvas lengths, and two long-canvas schedules for LLaDA-8B that separate model evaluations from commitments per step.

Canvas length. At the default canvas length, gating improves LLaDA-8B accuracy on MATH-500 by $+9.1$ pp over the pooled evaluation (95% CI [+5.6, +12.7]), with gains in both disjoint subsets (Table 6). As the canvas grows from 256 to 1,024 positions, the end-anchored share falls from 96.4% to 44.1% for LLaDA-8B and from 98.8% to 49.4% for LLaDA-1.5, and the gain from gating falls with it (Figure 3, Table 5). For LLaDA-1.5 the gain within end-anchored trajectories stays between $+7.3$ and $+13.6$ pp at every length, so its declining overall gain is compositional; for LLaDA-8B the gain within the group also shrinks (Table 1). Two long-canvas schedules that separate the number of model evaluations from the commitment density give the same small gains (Appendix B). Dream-7B is end-anchored on only 7.7–18.2% of problems; its one clear gain, at L=1024, comes from left-to-right responses that overrun the canvas (Appendix C).

Prompt. Explicit step-by-step prompting increases the advantage of gating on GSM8K by $+3.3$ pp (95% CI $[+1.1, +5.4]$; Table 2), and it does so by making end-anchoring more frequent. The end-

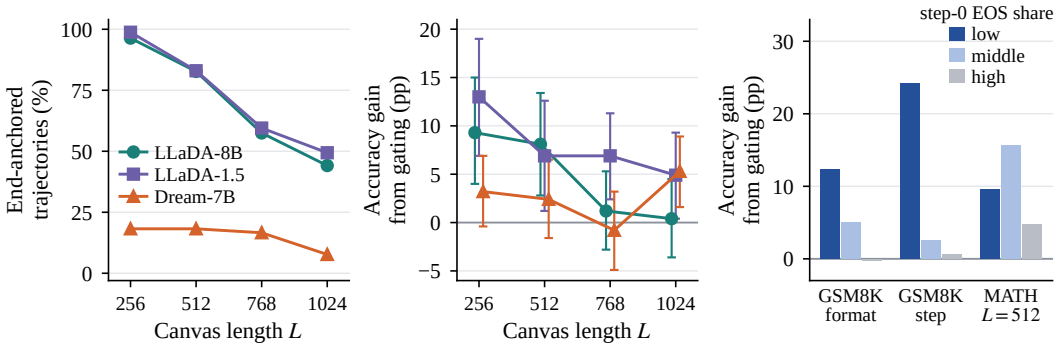

Figure 3: **End-anchoring, the canvas, and the first forward pass.** Left: share of end-anchored unrestricted trajectories on MATH-500 by canvas length. Middle: gated minus unrestricted accuracy on all scored items, with 95% paired-bootstrap intervals. Right: gain from gating by tercile of the step-0 EOS share, on held-out problems (GSM8K: remaining 1,069 questions under the format-only and step-by-step instructions; MATH: remaining 250 problems).

anchored share rises from 18.1% to 26.2%, while the gain within end-anchored trajectories (+32.0 and +33.6 pp) and outside them (−0.2 and +0.1 pp) is unchanged; share times within-group gain gives +5.8 and +8.8 pp, close to the observed overall gains of +5.6 and +8.9 pp. On MATH, where most trajectories are end-anchored under either prompt, the interaction is inconclusive (Appendix A).

Prediction before decoding. Whether a trajectory will be end-anchored can be read from the first forward pass. The step-0 EOS share is lower for trajectories that become end-anchored (AUC 0.83–0.94 across the LLaDA conditions in which end-anchoring is not nearly universal; Figure 9). A threshold chosen on the first 247 GSM8K questions classifies 87.5% of the remaining 1,069 correctly, and gating gains +27.6 pp on the 116 questions it flags against +2.9 pp on the rest (Figure 3, right; Table 9). The threshold depends on the prompt and the canvas length, and the signal is weak for Dream-7B (AUC 0.57–0.66; Appendix D).

3.3 Holding back the end removes the cost; delaying the answer does not

If end-anchoring is what costs accuracy, then preventing the end from being fixed early, and nothing else, should recover the gain of gating. We test this with a pre-specified replay on the remaining 1,069 GSM8K questions: from a completed unrestricted reference, the last 32 positions of the response are held ineligible until step 448 of 512 while everything else is decoded as before (*tail held*); the control holds an equally long contiguous span inside the central half of the response for the same time (*middle span held*).

Holding back the end. On the 194 questions that are end-anchored in the reference, the tail-held arm raises accuracy from 32.0% to 63.4%, the same as gating (63.9%), whereas the middle-span arm leaves it at 32.5% (Figure 4, right). The pre-specified contrast, tail held minus middle span held, is +30.9 pp (95% CI [+22.7, +39.2]; 71/11 discordant pairs, exact McNemar $p{=}6.8 \times 10^{-12}$), and gating minus tail held is +0.5 pp ([-3.6, +4.6]). The intervention removes end-anchoring rather than relocating it: no tail-held trajectory is end-anchored, none anchors on the positions just before the held region, and the response is written from left to right (median position–order correlation 1.00) and ends where the gated solution ends (median length 430 against 424), whereas the middle-span arm stays end-anchored on every question. On the 875 questions that are not end-anchored in the reference, no arm changes accuracy by more than about two points (Table 10).

Delaying the answer. Delaying only the answer position, by contrast, leaves the anchor in place. In a terminal-choice scaffold that locates the answer before generation without supplying its content (Section 2.4), delaying the answer improves LLaDA-8B accuracy over the matched reasoning delay by +1.6 pp on the full test set (95% CI [+0.5, +2.7]; Figure 4), with release at each item's reference reasoning-median step. Against a pre-specified control that delays a randomly drawn reasoning

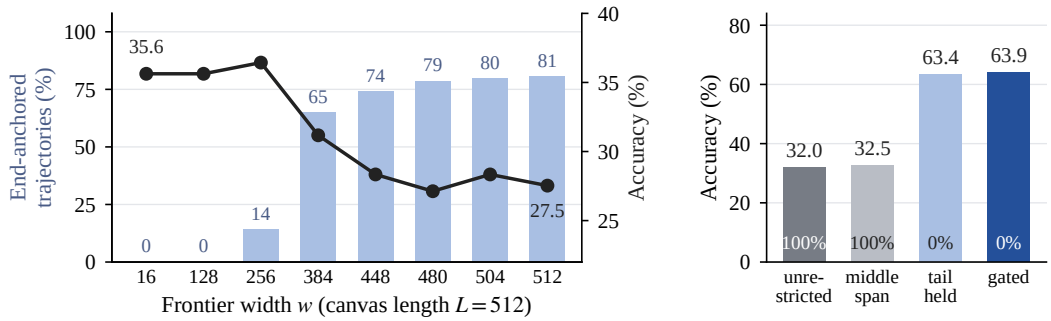

Figure 4: **Two pre-specified interventions on the anchor.** Left: frontier-width sweep on MATH-500 (L=512, 247 problems, one job array): share of end-anchored trajectories (bars) and accuracy (line); once the window reaches the end of the canvas, end-anchoring returns and accuracy falls. Right: tail-delay replay on the 194 GSM8K questions that are end-anchored in the unrestricted reference: 32 reference positions are held ineligible until step 448 of 512, either a contiguous span in the middle of the response or its tail; numbers inside the bars give the share of trajectories that are end-anchored.

position instead, the gain is $+2.0$ pp (95% CI [+0.8, +3.3]; exact McNemar p=0.001), and random-position delays leave accuracy at the level of free generation ($+0.1$ pp), so the effect is specific to the answer position but small (Appendix E). A longer delay does not increase the benefit, and the slot's own prediction of the answer is mostly settled early (Appendix E). Without the scaffold the effect is smaller still: delaying the answer positions of a completed unrestricted reference until step 256 of 512, against timing-matched reasoning positions, gains $+0.8$ pp over all 1,069 questions (95% CI $[+0.0, +1.7]$), against $+5.6$ pp for gating on the same questions; The delay leaves the anchor in place: of the 136 active questions that are end-anchored in the reference, 90% remain end-anchored when the answer is delayed, and their accuracy rises from 31.6% to 38.2%, against 64.7% under gating (Appendix E).

Releasing the restriction. The converse holds as well: the gain from gating should vanish once the window is wide enough for the end of the canvas to be committed early. A pre-specified sweep of the frontier width up to the canvas length confirms this (MATH-500, L=512, all arms in one job array; Figure 4, left). For $w \leq 128$ no trajectory is end-anchored and accuracy is 35.6%; at w=384 and w=512 the end-anchored share returns to 65% and 81%, and accuracy falls to 31.2% and 27.5% (-8.1 pp against w=16, 95% CI [-13.4, -2.8]). The pre-specified contrast, within problems, between runs in which a problem is end-anchored and runs in which it is not is -11.0 pp (95% CI [-17.5, -4.7]; 165 problems), and the loss is confined to problems that become end-anchored (Table 8).

4 Related work

Answer exposure. Yu et al. (2026a) study requested answer/reasoning order, complete-answer exposure, sampling order, and generation-length failures on LLaDA. We locate the accuracy cost in end-anchored trajectories, relate it to the prompt and the canvas, and intervene on the anchor itself; complete-answer exposure and irreversible commitment coincide in our monotone runs.

Order restrictions. Block, set, and windowed diffusion restrict order for speed or quality (Arriola et al., 2025; Arriola & Kuleshov, 2026; Zuo et al., 2026). Ni et al. (2026) show how confidence-based sampling postpones uncertain logical choices, reducing solution coverage; end-anchoring describes which region of the response is fixed first, and both can coexist. Deferred Commitment Decoding (Shu et al., 2026) adapts a window anchored at the leftmost mask using confidence thresholds, whereas our fixed-width diagnostic keeps the released within-window ranking and transfer schedule.

Commitment policies. Decoders use answer convergence (Li et al., 2025), latent belief mixtures (Zhu et al., 2025b), temporal stability (Wu & Huang, 2026; Wang et al., 2026a), learned commitment

(Sun et al., 2026), block closure (Wang et al., 2026b), or candidate-aware stopping (Lee et al., 2026). We analyze commitment order under fixed policies (cf. Zhou et al., 2026).

Revision and termination. Remasking revises earlier predictions (Wang et al., 2025; Yu et al., 2026b), whereas we delay positions before they commit. EOS overflow and variable-length generation concern termination (Kim et al., 2025; Liu et al., 2026; Yang et al., 2025; 2026); our primary comparisons keep EOS/EOT handling fixed.

5 Discussion

On both benchmarks the gain from frontier gating lies in end-anchored trajectories, in which the closing sentence and then the answer are fixed at the end of the canvas before the reasoning exists. Answer-first commitment and canvas filling are its visible signatures; prompt wording, problem difficulty, and canvas length change how often it occurs, not what it costs, so prompt effects need evaluation under the decoder that realizes them. Three interventions bracket this reading. Holding back only the tail of the response removes end-anchoring and recovers the whole gain of gating; widening the gate until the end of the canvas becomes available restores end-anchoring and removes the gain; and delaying the answer alone leaves the anchor in place and gains little (Figure 4). The benefit therefore comes from constructing the reasoning left to right instead of filling it in behind a fixed ending (Ni et al., 2026), and it is available without restricting the order of the rest of the response.

End-anchoring can be anticipated from the step-0 EOS share before decoding, and a left-to-right restriction is harmless on the other trajectories, so the signal matters more for understanding when unrestricted decoding fails, and for choosing a canvas length, than as a switch between decoders.

6 Limitations

GSM8K prompt and delay experiments use LLaDA-8B; MATH canvas experiments include other models without controlling training differences, and schedule controls do not isolate canvas effects (Appendix B). The end-anchoring definition was chosen after inspecting the commitment traces, and the split by end-anchoring is observational: it is a property of the unrestricted trajectory and more frequent on harder problems. The tail-delay replay and the width sweep were specified in advance, but the replay holds positions defined by a completed unrestricted run, so it is a diagnostic rather than a decoding method, and the sweep manipulates end-anchoring only through the window; the step-0 predictor uses condition-specific thresholds and is weak for Dream-7B (Appendix C). Experiments cover two mathematical benchmarks at temperature zero with at most two commitments per step; denser parallel schedules, sampling, other task families, and block-diffusion models are outside our scope. Primary results condition on EOS/EOT suppression (Appendix G). Answer-span mapping, which affects answer-first status but not end-anchoring, was developed on the human-reviewed sample (Appendix F).

The answer-delay effect is measured in a terminal-choice scaffold that fixes the answer's location and length, and its controls do not equate the semantic information that is delayed; the free-form replays follow a reference generation, so they evaluate neither an online policy nor a semantic answer delay.

7 Conclusion

Displayed reasoning order does not reveal when an answer becomes irreversible. In LLaDA models the cost of unrestricted commitment order lies in end-anchored trajectories, which fix the end of the response, answer included, before its middle. They can be anticipated from the first forward pass and vary with the prompt and the canvas; holding back only the tail of the response removes them together with the accuracy loss, delaying the answer does not, and widening a left-to-right restriction to the canvas length brings both back.

AI USE STATEMENT

In this work, we used generative AI tools (large language model assistants) for the following tasks that require disclosure. *Proposing and refining hypotheses and the conceptual framing:* the end-anchoring analysis and the use of the step-0 EOS share as a predictor were suggested by an AI assistant while it reviewed an earlier draft together with the per-item records, and were then evaluated on the full set of commitment traces. *Feedback on methodology and experiments:* AI-written mock reviews of drafts identified missing controls and suggested additional conditions, some of which we ran. *Implementation:* AI assistants helped write parts of the decoding and logging code, the scripts that derive per-item records from the commitment traces, and the analysis scripts. *Interpreting results:* interpretations proposed by AI assistants were accepted, revised, or rejected by the authors after checking them against the data. We did not use generative AI tools to generate data, to formulate or prove mathematical claims, or for translation; qualitative data analysis is not applicable to this work. Additionally, we used generative AI tools to write and edit plotting code and figures, to draft and edit text in every section, to restructure the paper and its appendix, and to propose keywords.

We have reviewed all AI-assisted work. AI-written analysis code was checked by reproducing quantities established independently beforehand (for example, the numbers of correct, answer-first, and canvas-filling responses on the GSM8K evaluation) and ships with read-only consistency checks. Definitions that depend on the decoding engine (the step-0 EOS share, response length, and the terminator tokens of each model) and all statements about hardware and sampler builds were verified by the authors against the source code and run logs; one AI-produced claim about Dream-7B that rested on a misread log column was corrected in this way. All experiments were run by the authors. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Marianne Arriola and Volodymyr Kuleshov. Set diffusion: Interpolating token orderings between autoregression and diffusion for fast and flexible decoding. In *International Conference on Machine Learning*, 2026.

Marianne Arriola, Aaron Gokaslan, Justin T Chiu, Zhihan Yang, Zhixuan Qi, Jiaqi Han, Subham Sekhar Sahoo, and Volodymyr Kuleshov. Block diffusion: Interpolating between autoregressive and diffusion language models. In *International Conference on Learning Representations*, 2025.

Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems*, 2021.

Karl Cobbe et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Dan Hendrycks et al. Measuring mathematical problem solving with the math dataset. In *NeurIPS Datasets and Benchmarks*, 2021.

Bumjun Kim, Dongjae Jeon, Dueun Kim, Wonje Jeung, and Albert No. Rainbow padding: Mitigating early termination in instruction-tuned diffusion LLMs. *arXiv preprint arXiv:2510.03680*, 2025.

Tamera Lanham et al. Measuring faithfulness in chain-of-thought reasoning. *arXiv preprint arXiv:2307.13702*, 2023.

Chia-Ming Lee, Ming-Ching Chang, Xin Li, Yu-Lun Liu, and Chih-Chung Hsu. Where and when to commit: Candidate-aware decoding for diffusion language models. *arXiv preprint arXiv:2607.28166*, 2026.

Pengxiang Li, Yefan Zhou, Dilxat Muhtar, Lu Yin, Shilin Yan, Li Shen, Soroush Vosoughi, and Shiwei Liu. Diffusion language models know the answer before decoding. *arXiv preprint arXiv:2508.19982*, 2025.

Hunter Lightman et al. Let's verify step by step. In *International Conference on Learning Representations*, 2024. MATH-500 subset origin.

Chunyu Liu, Zhengyang Fan, Kaisen Yang, and Alex Lamb. VoidPadding: Let [VOID] handle padding in masked diffusion language models so that [EOS] can focus on semantic termination. *arXiv preprint arXiv:2606.17999*, 2026.

Zanlin Ni, Shenzhi Wang, Yang Yue, Tianyu Yu, Weilin Zhao, Yeguo Hua, Tianyi Chen, Jun Song, Cheng Yu, Bo Zheng, and Gao Huang. The flexibility trap: Rethinking the value of arbitrary order in diffusion language models. *arXiv preprint arXiv:2601.15165*, 2026. ICML 2026.

Shen Nie, Fengqi Zhu, Zebin You, Xiaolu Zhang, Jingyang Ou, Jun Hu, Jun Zhou, Yankai Lin, Ji-Rong Wen, and Chongxuan Li. Large language diffusion models. *arXiv preprint arXiv:2502.09992*, 2025.

Yingte Shu, Yuchuan Tian, Chao Xu, Yunhe Wang, and Hanting Chen. Deferred commitment decoding for diffusion language models. *arXiv preprint arXiv:2601.02076*, 2026.

Bohang Sun, Max Zhu, Francesco Caso, Jindong Gu, Junchi Yu, Philip Torr, Pietro Liò, and Jialin Yu. The path matters: Learning a token-commitment policy for diffusion language models. *arXiv preprint arXiv:2605.24697*, 2026.

Miles Turpin, Julian Michael, Ethan Perez, and Samuel R Bowman. Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. In *Advances in Neural Information Processing Systems*, 2023.

Chengcheng Wang, Tingzhang Luo, Wenhao Li, Jianyuan Guo, and Chang Xu. TACG: Trajectory-aware commit gating for diffusion language model decoding. *arXiv preprint arXiv:2607.03236*, 2026a.

Danny Wang, Ruihong Qiu, and Zi Huang. When to commit? towards variable-size self-contained blocks for discrete diffusion language models. *arXiv preprint arXiv:2604.23994*, 2026b.

Guanghan Wang, Yair Schiff, Subham Sekhar Sahoo, and Volodymyr Kuleshov. Remasking discrete diffusion models with inference-time scaling. In *Advances in Neural Information Processing Systems*, 2025.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, 2022.

Yue Wu and Jian Huang. Stability-weighted decoding for diffusion language models. *arXiv preprint arXiv:2604.17068*, 2026.

Jingyi Yang, Yuxian Jiang, and Jing Shao. ρ-EOS: Training-free bidirectional variable-length control for masked diffusion LLMs. *arXiv preprint arXiv:2601.22527*, 2026.

Yicun Yang, Cong Wang, Shaobo Wang, Zichen Wen, Biqing Qi, Hanlin Xu, and Linfeng Zhang. Diffusion LLM with native variable generation lengths: Let [EOS] lead the way. *arXiv preprint arXiv:2510.24605*, 2025.

Jiacheng Ye, Zhihui Xie, Lin Zheng, Jiahui Gao, Zirui Wu, Xin Jiang, Zhenguo Li, and Lingpeng Kong. Dream 7B: Diffusion large language models. *arXiv preprint arXiv:2508.15487*, 2025.

Longxuan Yu, Yu Fu, Shaorong Zhang, Hui Liu, Mukund Varma T, Greg Ver Steeg, and Yue Dong. Thinking out of order: When output order stops reflecting reasoning order in diffusion language models. *arXiv preprint arXiv:2601.22035*, 2026a.

Longxuan Yu, Shaorong Zhang, Yu Fu, Hui Liu, Yue Dong, and Greg Ver Steeg. Revise, don't freeze: Sampler-matched training for self-correcting masked diffusion language models. *arXiv preprint arXiv:2606.01026*, 2026b.

Fan Zhou, Weitian Wang, and Tim Van de Cruys. Commitment before realization: When classifier-free guidance becomes unnecessary in masked diffusion language models. *arXiv preprint arXiv:2608.08082*, 2026.

Fengqi Zhu, Rongzhen Wang, Shen Nie, Xiaolu Zhang, Chunwei Wu, Jun Hu, Jun Zhou, Jianfei Chen, Yankai Lin, Ji-Rong Wen, and Chongxuan Li. LLaDA 1.5: Variance-reduced preference optimization for large language diffusion models. *arXiv preprint arXiv:2505.19223*, 2025a.

Qinglin Zhu, Yizhen Yao, Runcong Zhao, Yanzheng Xiang, Amrutha Saseendran, Chen Jin, Philip Teare, Bin Liang, Yulan He, and Lin Gui. Latent refinement decoding: Enhancing diffusion-based language models by refining belief states. *arXiv preprint arXiv:2510.11052*, 2025b.

Fengrui Zuo, Zhiwei Ke, Yiming Liu, Wenqi Lou, Chao Wang, and Xuehai Zhou. Window-Diffusion: Accelerating diffusion language model inference with windowed token pruning and caching. *arXiv preprint arXiv:2601.20332*, 2026.

A PROMPTS AND GSM8K ACCURACY

The three GSM8K instructions are reproduced below.

> **`Direct answer.`** `Respond with only the final answer in the form ``The answer is X''. Do not include any reasoning, explanation, or intermediate steps.`
>
> **`Format-only.`** `Finish your response with ``The answer is X'', where X is the final answer.`
>
> **`Explicit step-by-step.`** `Let's think step by step. Finish your response with ``The answer is X'', where X is the final answer.`

Under unrestricted decoding the format-only instruction already elicits reasoning (at least 20 tokens before the answer) in 95.9% of the remaining 1,069 responses, with a median length of 290 tokens, against 100% and 298 tokens under the explicit instruction; direct-answer responses have a median of 7 tokens. Every response has an extractable answer.

Table 2: GSM8K accuracy of LLaDA-8B by instruction and decoding policy. The first subset excludes three interface-development questions.

Questions	Instruction	Unrestricted	Gated (w=16)	Semi-AR (block 32)
First 247	Direct answer	0.457	0.457	0.457
	Format-only	0.777	0.846	0.842
	Explicit step-by-step	0.729	0.874	0.854
Remaining 1069	Format-only	0.769	0.825	–
	Explicit step-by-step	0.743	0.832	–

The prompt–policy interaction is $+7.7$ pp on the first subset (95% CI $[+2.8, +12.6]$; $+6.1$ pp with semi-autoregressive decoding in place of gating) and $+3.3$ pp on the remaining questions (95% CI $[+1.1, +5.4]$). On MATH-500 it is -1.2 pp (95% CI [-7.3, +4.9]) on the first 247 problems, $+1.6$ pp ([-3.6, +6.8]) on the remaining 250, and $+0.2$ pp ([-3.8, +4.0]) pooled.

Direct-answer check. All 247 direct-answer outputs are token-identical across the three policies, and their 512-position commitment-step arrays are identical (247 of 247). The largest committed content position is 11, inside the 16-position frontier for every response, so the restriction never binds.

B STATISTICS, HARDWARE, AND SCHEDULES

For paired binary-accuracy contrasts with sample size n and discordant counts b, c, we compute percentile intervals using 100,000 draws with seed 20260906 from $\mathrm{Multinomial}(n; b/n, c/n, 1-(b+c)/n)$ and report $100(b^* - c^*)/n$; this is equivalent in distribution to resampling questions. The exact two-sided McNemar p-value uses $b \sim \mathrm{Binomial}(b+c, 1/2)$ under the null. Multi-condition interactions resample each question's four outcomes jointly. Group-wise gains (end-anchored versus other trajectories, strata, signatures) draw question types from one multinomial over all questions of a condition, so group sizes vary across draws as they do under question resampling.

Hardware and sampler builds. Within every paired unrestricted-versus-gated (and blockwise) comparison in the tables, both arms were generated by the same sampler build in the same job array on a single GPU class (A6000 or RTX 3090), never mixed within a pair. For the slot interventions, the free reference and the delay arms share the GPU class and the engine revision but were run as separate jobs, and the declared-slot experiments used one GPU class per model. Two exceptions apply. The release-sweep arms at $r \in \{70, 210\}$ were run on RTX 3090 against a free reference generated on A6000, after a 40-item replay had been verified to be token-exact across the two classes. The diagnostics that reuse earlier frozen cells (the suppression-off cells of Appendix G) were generated on A100 hardware, where the same code changes about 34% of tokens relative to A6000; these cells are never paired against cells generated elsewhere. The three pre-specified interventions (Appendices D

and E) were each generated as one job array on one GPU class, the width sweep on RTX 3090 and the tail-delay and induction replays on A6000, with all arms of an experiment, including its unrestricted reference and its gated arm, in the same array.

Schedules. In the canvas-length study the number of denoising steps equals L, with one commitment per step; realized commitments per step equal the nominal value in every condition. At $L{=}1024$ a second schedule uses 512 steps with two commitments per step, which retains the number of model evaluations of the $L{=}512$ condition, whereas the 1,024-step schedule retains the commitment density. Both give small gating gains ($+0.4$ pp, 95% CI [-3.6, +4.5], and $+0.8$ pp, 95% CI [-4.0, +5.7]; Table 5); the intervals do not establish equivalence, and matching evaluations does not match total computation.

C END-ANCHORING

Definition and robustness. The response is the content before the first terminator token (end of sequence or end of turn) on the canvas, under the same rule for all three models; in Dream-7B's free-form outputs that token is always <|endoftext|>. A response is end-anchored when the median commitment step of its last 16 tokens precedes that of the central half of the response; responses shorter than 40 tokens are not end-anchored. Status is read from the unrestricted trajectory before the gated outcome is examined. The difference between the two medians is bimodal with almost no mass near zero (Figure 5), so the classification does not depend on a threshold: using the last 8 or 32 tokens, or all earlier positions as the comparator, changes the end-anchored share only between 17.2% and 18.7% on GSM8K and between 81.7% and 82.9% on MATH ($L{=}512$), with gains inside of $+31.0$ to $+32.5$ pp and $+10.7$ to $+10.8$ pp. Figure 6 shows one trajectory of each kind.

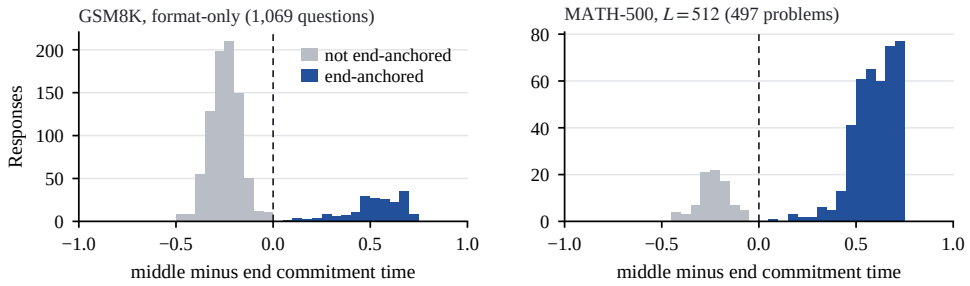

Figure 5: **End-anchoring is a dichotomy, not a threshold.** Median commitment time of the central half of the response minus that of its last 16 tokens, for unrestricted LLaDA-8B responses. Positive values are end-anchored.

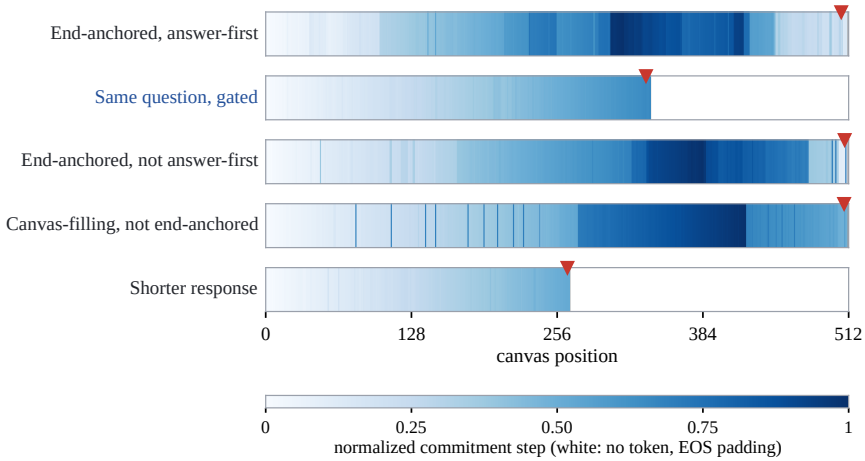

Figure 6: **Example trajectories** (GSM8K, format-only). Each strip shows when every canvas position commits; the red marker is the mapped answer. Examples are the first question, in index order, of each type, and the second strip is the gated run of the first question. End-anchored responses fix their final tokens early and fill the middle last; the other responses are written from left to right.

All conditions. Figure 7 gives the end-anchored share and the gain from gating inside and outside end-anchored trajectories for every condition. In the LLaDA conditions every answer-first trajectory but two is end-anchored (139 of 139 on GSM8K; 351 of 351 on MATH at L=512), every end-anchored response is canvas-filling, no gated or blockwise trajectory is end-anchored, and trajectories that are not end-anchored are committed from left to right (median correlation between position and commitment order at least 0.994 wherever ten or more such trajectories exist). Dream-7B end-anchors on 3–18% of problems; its canvas-filling responses are mostly left-to-right overruns.

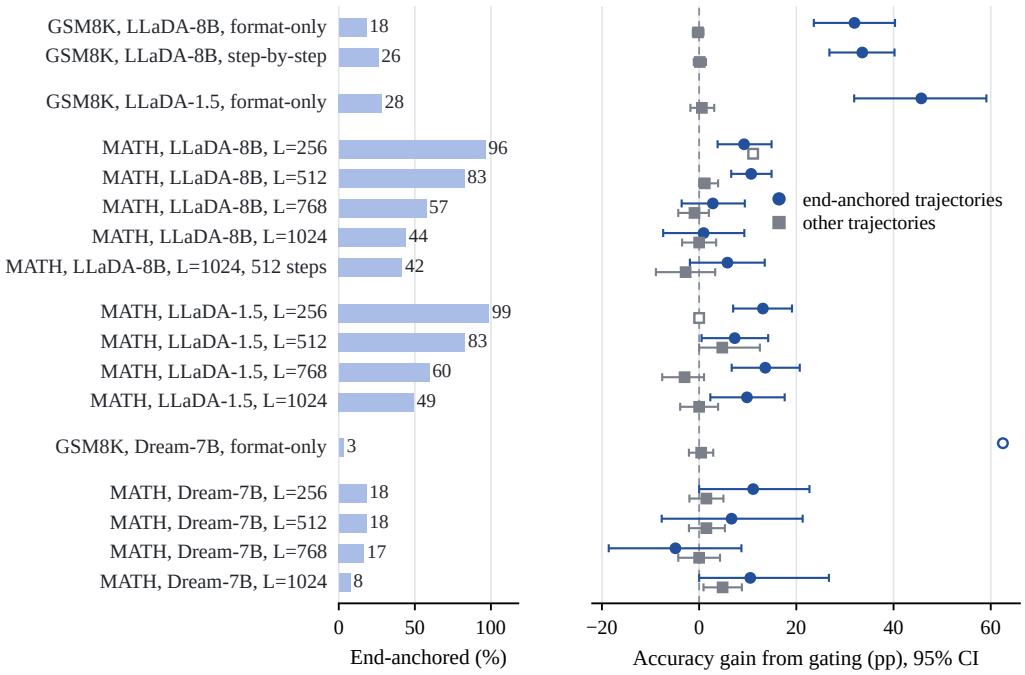

Figure 7: **End-anchoring in every condition** (format-only instruction unless stated). Left: share of unrestricted trajectories that are end-anchored. Right: gated minus unrestricted accuracy inside and outside end-anchored trajectories with 95% intervals; hollow markers are groups with fewer than ten problems, shown without an interval. GSM8K rows use the remaining 1,069 questions for LLaDA-8B and the first 247 otherwise; MATH rows use 247 problems, except $L{=}512$ for LLaDA-8B (pooled 497).

Table 3: Which property of the unrestricted trajectory carries the gain. End-anchored trajectories gain whether or not they are answer-first (those that are not have their closing sentence fixed early but their answer value committed later); canvas-filling responses that are not end-anchored show no clear gain. AF: answer-first. GSM8K: format-only instruction; MATH: pooled 497 problems at $L{=}512$.

Condition	Trajectory	n	Accuracy, unres. → gated	Gain (pp, 95% CI)
GSM8K	end-anchored, AF	139	32.4% → 65.5%	+33.1 [+23.3, +42.8]
	end-anchored, not AF	55	30.9% → 60.0%	+29.1 [+12.3, +45.5]
	canvas-filling only	29	58.6% → 62.1%	+3.4 [-17.2, +24.1]
	shorter	846	87.8% → 87.5%	−0.4 [-1.3, +0.5]
MATH	end-anchored, AF	351	16.2% → 26.2%	+10.0 [+5.5, +14.5]
	end-anchored, not AF	60	21.7% → 36.7%	+15.0 [+3.1, +27.4]
	canvas-filling only	6	16.7% → 33.3%	+16.7 [+0.0, +50.0]
	shorter	80	78.8% → 78.8%	+0.0 [+0.0, +0.0]

Support for the answer when it commits. An instrumented replay of the 1,069 open-ended GSM8K questions records the model's predictions at the answer positions while they are still masked. Table 4 reports, at the last step before any answer token commits, how often the top prediction equals the token eventually committed and the mean log-probability of that token, together with the step at which the first answer token commits. In end-anchored trajectories the answer starts to commit early and with weak support, whether or not the trajectory is answer-first; under gating the same questions commit the answer late and with support close to that of the other trajectories. These are retrospective diagnostics of the committed tokens, not online confidence estimates.

Table 4: Support for the answer tokens at the last step before any of them commits (GSM8K, format-only, remaining 1,069 questions; groups defined on the unrestricted trajectory). Top-1: share of answer positions at which the model's most likely token, one step before the answer starts to commit, equals the token that ends up there (a low value means that the answer was still changing when it began to be fixed); $\log p$: mean log-probability of that token; first step: mean step, of 512, at which the first answer token commits.

		Unrestricted			Gated, same questions		
Unrestricted trajectory	n	top-1	$\log p$	first step	top-1	$\log p$	first step
End-anchored	194	71%	−0.97	103	94%	−0.61	419
answer-first	139	74%	−0.92	80	96%	−0.57	421
not answer-first	55	64%	−1.11	163	91%	−0.69	415
Other trajectories	875	97%	−0.24	255	98%	−0.19	260

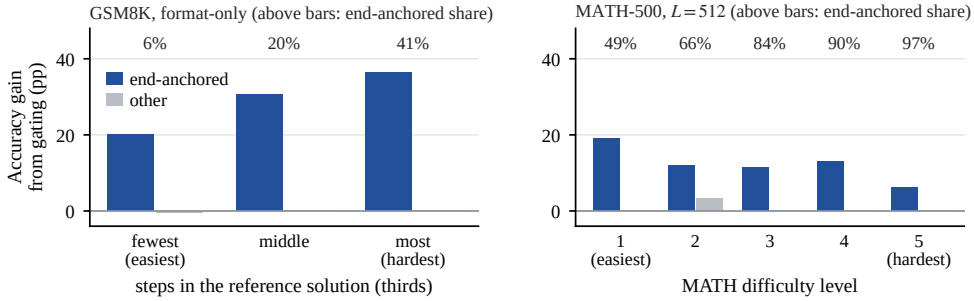

Figure 8: **The split is not a difficulty effect.** Gain from gating inside and outside end-anchored trajectories within difficulty strata (LLaDA-8B); percentages above the bars give the end-anchored share of each stratum. GSM8K (remaining 1,069 questions) has no difficulty labels, so questions are ordered by the number of steps in their reference solution and split into thirds; MATH (pooled 497 problems) uses its five difficulty levels. End-anchoring is more frequent on harder problems, but within every stratum the gain lies in end-anchored trajectories.

D CANVAS LENGTH, FRONTIER WIDTH, AND PREDICTION

Table 5: MATH-500 across models and canvas lengths (first 247 problems, format-only instruction; steps equal L except ‡512 steps with two commitments per step). Answer-first rate and median backfill are taken among outputs with a complete mapped answer span. Gated decoding matches or exceeds unrestricted decoding with twice the canvas in every point comparison.

		Accuracy			Unrestricted trajectories		
Model	L	unres.	gated	Gain (pp, 95% CI)	End-anch.	Answer-first	Backfill
LLaDA-8B	256	0.202	0.296	+9.3 [+4.0, +15.0]	96%	83%	78%
	512	0.275	0.356	+8.1 [+2.8, +13.4]	81%	68%	72%
	768	0.328	0.340	+1.2 [-2.8, +5.3]	57%	45%	41%
	1024	0.348	0.352	+0.4 [-3.6, +4.5]	44%	30%	1%
	1024‡	0.336	0.344	+0.8 [-4.0, +5.7]	42%	33%	1%
LLaDA-1.5	256	0.206	0.336	+13.0 [+6.9, +19.0]	99%	90%	82%
	512	0.267	0.336	+6.9 [+1.2, +12.6]	83%	73%	77%
	768	0.291	0.360	+6.9 [+2.4, +11.3]	60%	54%	63%
	1024	0.336	0.385	+4.9 [+0.4, +9.3]	49%	41%	4%
Dream-7B	256	0.385	0.417	+3.2 [-0.4, +6.9]	18%	12%	2%
	512	0.405	0.429	+2.4 [-1.6, +6.5]	18%	16%	1%
	768	0.405	0.397	−0.8 [-4.9, +3.2]	17%	14%	1%
	1024	0.372	0.425	+5.3 [+1.6, +8.9]	8%	7%	1%

Pooled short-canvas accuracy. The two disjoint format-only LLaDA-8B evaluations at $L{=}512$ give 68 versus 88 correct on the first 247 problems and 66 versus 91 on the remaining 250. The pooled gain is $+9.1$ pp (95% CI [+5.6, +12.7]; discordant pairs 64/19; exact McNemar $p{=}7.4\times10^{-7}$; Table 6); it pools two disjoint runs and is not an additional independent evaluation.

Table 6: LLaDA-8B format-only accuracy at $L{=}512$ from the per-question records.

Subset	n	Unrestricted	Gated	Gain (pp)	95% CI	b/c	Exact McNemar p
First 247	247	0.275	0.356	$+8.1$	[+2.8, +13.4]	32/12	0.004
Remaining 250	250	0.264	0.364	$+10.0$	[+5.2, +14.8]	32/7	7.0×10^{-5}
Pooled 497	497	0.270	0.360	$+9.1$	[+5.6, +12.7]	64/19	7.4×10^{-7}

Frontier width. Every finite frontier width exceeds unrestricted accuracy, with no clear trend across widths (Table 7), and none of them produces an end-anchored trajectory: 80.6% of the 247 unrestricted trajectories are end-anchored, against none for $w\in\{1,4,16,64,128\}$ and for blockwise decoding. The $w{=}16$ and unrestricted conditions reuse the main generations.

Table 7: LLaDA-8B frontier-width sweep ($L{=}512$, $n{=}247$ per condition), scored with the normalized rule used throughout.

Policy	Accuracy
Gated $w{=}1$	0.336
Gated $w{=}4$	0.356
Gated $w{=}16$	0.356
Gated $w{=}64$	0.344
Gated $w{=}128$	0.356
Unrestricted	0.275
Semi-AR (block 32)†	0.385

Frontier width up to the canvas length. A second sweep, with the analysis plan fixed before generation and all eight arms run in one job array, widens the frontier until it covers the whole canvas (Table 8); the arm $w{=}512$ is token-identical to the unrestricted code path on 20 of 20 checked problems. The primary contrast is taken within problems: among the 165 problems whose end-anchored status differs across the runs with $w\geq256$, accuracy in the end-anchored runs minus accuracy in the other runs is -11.0 pp (95% CI [-17.5, -4.7]; item bootstrap). Most of these problems are end-anchored in five of the six runs, so the comparison leans on $w{=}256$, and two exploratory analyses separate end-anchoring from the width itself. With $w{=}16$ as a common reference, the four widest windows lose -9.4 pp ([-15.5, -3.5]) on problems that are end-anchored at that width and -1.9 pp ([-6.3, +1.0]) on the others. Among problems that are not end-anchored at $w{=}256$, the 125 that become end-anchored at $w{=}384$ lose -8.8 pp between the two widths while the 87 that do not change by $+0.0$ pp (difference in differences -8.8 pp, 95% CI [-16.0, -1.6]). End-anchored trajectories first commit a token among their last 16 positions a median of 4–30 steps after those positions become eligible.

Table 8: Pre-specified frontier-width sweep up to the canvas length (LLaDA-8B, MATH-500, L=512, 247 problems, one job array). Differences are against w=16 on the same problems, with intervals as in Appendix B; the last two columns split each arm by whether the trajectory is end-anchored at that width.

					End-anchored at w / other	
w	Accuracy	vs. w=16 (pp, 95% CI)	End-anch.	Canvas-fill.	Accuracy	vs. w=16 (pp)
16	35.6%	–	0%	53%	– / 35.6%	– / –
128	35.6%	+0.0 [-3.2, +3.2]	0%	59%	– / 35.6%	– / +0.0
256	36.4%	+0.8 [-3.2, +4.9]	14%	67%	31.4% / 37.3%	+5.7 / +0.0
384	31.2%	−4.5 [-8.9, -0.4]	65%	75%	18.1% / 55.2%	−5.6 / −2.3
448	28.3%	−7.3 [-12.1, -2.4]	74%	79%	16.9% / 60.9%	−9.3 / −1.6
480	27.1%	−8.5 [-13.8, -3.6]	79%	81%	17.0% / 64.2%	−10.3 / −1.9
504	28.3%	−7.3 [-12.6, -2.4]	80%	82%	18.8% / 66.0%	−8.6 / −2.0
512	27.5%	−8.1 [-13.4, -2.8]	81%	82%	17.6% / 68.8%	−9.5 / −2.1

Prediction from the first forward pass. At every decoding step the engine records the fraction of still-masked positions whose full-vocabulary argmax, taken from the raw logits before EOS/EOT confidence suppression, is a terminator token. For the LLaDA models these are the padding token <|endoftext|> and the end-of-turn token <|eot_id|>, which occupies a single position; for Dream-7B they are <|endoftext|>, which it predicts on 74–93% of the empty canvas, and <|im_end|>, which it never predicts there. At step 0 all L positions are masked, so the step-0 value is the share of the canvas that would become a terminator if every position were committed at once. Figure 9 and Table 9 evaluate it on held-out problems. The threshold depends on the prompt and the canvas length, so it has to be set per condition; on MATH at L=512, where most trajectories are end-anchored, classification accuracy barely exceeds the base rate although the ranking remains informative.

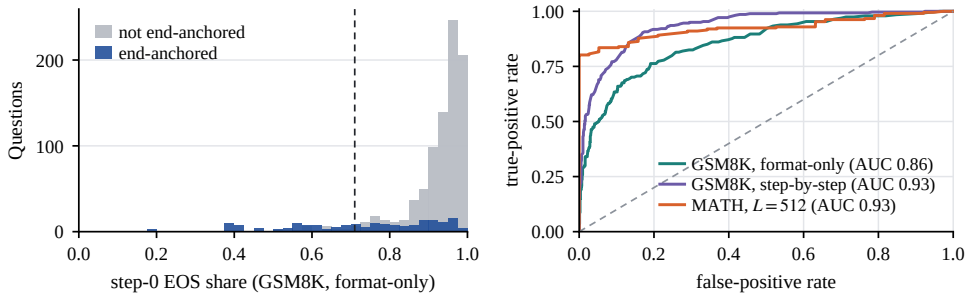

Figure 9: **The step-0 EOS share separates end-anchored trajectories.** Left: distribution on the remaining 1,069 GSM8K questions; the dashed line is the threshold chosen on the first 247. Right: ROC curves on held-out problems (GSM8K: remaining 1,069 questions; MATH: remaining 250 problems).

Table 9: Step-0 EOS share as a predictor of end-anchoring. The threshold is chosen on one set of problems and applied to a disjoint set; "flagged" are test problems below it.

	Problems		Test	Base		Gating gain (pp)	
Condition	train → test	Threshold	accuracy	rate	AUC	flagged	not flagged
GSM8K, format-only	247 → 1,069	0.71	87.5%	18.1%	0.86	+27.6 (n=116)	+2.9
GSM8K, step-by-step	247 → 1,069	0.46	86.6%	26.2%	0.93	+33.5 (n=179)	+3.9
MATH, L=512	247 → 250	0.65	85.2%	84.8%	0.93	+10.3 (n=223)	+7.4

E DELAY EXPERIMENTS

Construction. Each GSM8K question is converted into a four-option problem. Three distinct distractors are drawn from training-set answers with the same order of magnitude as the correct value, and the options are assigned to A–D with a fixed per-question permutation; the labels are one tokenizer token, and the majority-label baseline is 26.7%. The scaffold contains 280 masked reasoning positions and one masked answer position; delimiters and the termination tail are fixed before generation. A free reference pass records commitment times. The matched reasoning delay holds back the reasoning token that commits before release and is closest in time to the answer (median and 90th-percentile timing error: one step); selection uses commitment times only. The declared slot commits at a median of 6 of 281 steps for LLaDA-8B, far earlier than in ordinary generation, so the scaffold is an extreme answer-first regime.

All contrasts. Figure 10 collects every answer-minus-control contrast. The unlabeled slot was evaluated on a 679-question sample and then on the 640-question complement ($+2.7$ pp and $+0.5$ pp; difference $+2.2$ pp, 95% CI $[-0.1, +4.5]$); the main text reports the pooled estimate. Labeling the slot with "Final answer:" does not increase the effect (declared minus unlabeled on the same 679 questions: -1.3 pp, 95% CI $[-3.4, +0.7]$). Releases at steps 70, 140, and 210 of 281 give similar contrasts (Holm-adjusted p=0.026, 0.026, and 0.026); every between-release interval includes zero, the answer is held for a median of 66, 138, and 209 additional steps, and valid outputs remain at 100.0–100.0%. Dream-7B is off the scale of Figure 10, but its slot interface itself degrades the model: it answers 0.767 of 679 questions correctly in free-form generation and 0.571 under the declared slot, so its delay gain is recovery within a degraded interface, and the main text reports delay experiments for LLaDA-8B only.

Random-position controls. A later set of runs added two controls on the same 1,319 questions, with the analysis plan, including the primary contrast, fixed before generation. One control delays a randomly drawn reasoning position among those that commit before release in the free reference; the other draws the position uniformly. The primary contrast, answer delay minus the random active position, is $+2.0$ pp (95% CI [+0.8, +3.3]; discordant pairs 47/20; exact McNemar p=0.001). The remaining contrasts are exploratory, with Holm adjustment within the family: answer delay minus the uniform position, $+1.9$ pp ([+0.7, +3.1]; Holm p=0.021), and minus the timing-matched position, $+1.6$ pp ([+0.5, +2.7]; Holm p=0.052). Neither random control differs from free generation ($+0.1$ pp, [-0.6, +0.8], and $+0.2$ pp, [-0.4, +0.8]), nor does the timing-matched control ($+0.5$ pp, [-0.2, +1.3]). Repeating the random draw five times on 256 questions changes the outcome of 8 of them, so a single draw is a stable comparator.

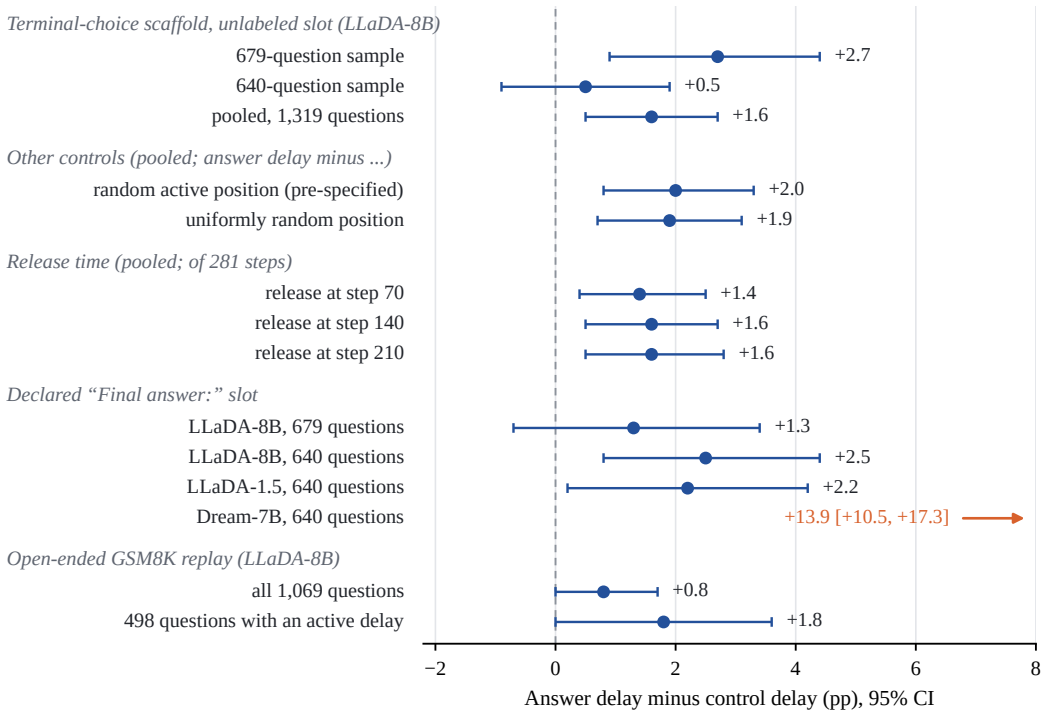

Figure 10: **Every answer-delay contrast.** Accuracy under answer delay minus accuracy under the matched reasoning delay, with 95% paired-bootstrap intervals. The release-time rows and the pooled row use the same 1,319 questions; step 140 is the reference reasoning-median release.

Answer-slot dynamics. Instrumented replays record what the masked answer slot predicts at every step (Figure 11). Free generation commits the slot almost immediately: the median commitment step is 7 and the mean 21 of 281, when the slot gives the gold option a probability of 0.74 with an entropy of 0.51 nats over the four options. When the slot is instead held masked, all 1,319 questions remain observable until release, and the probability of the gold option rises from 0.54 at step 0 to 0.86 while the entropy falls to about 0.1 nats; the delayed slot then commits at 0.87 and 0.09 nats. The reasoning written in the meantime therefore sharpens a preference that is mostly present early: the option finally chosen differs from that of free generation on 5.7% of questions under answer delay and 2.1% under the timing-matched delay. The flat, uncertain curves of the other arms are a selection effect, because confidence-ordered decoding removes a slot from the average as soon as it becomes confident, and the same applies to the answer-delay arm after release.

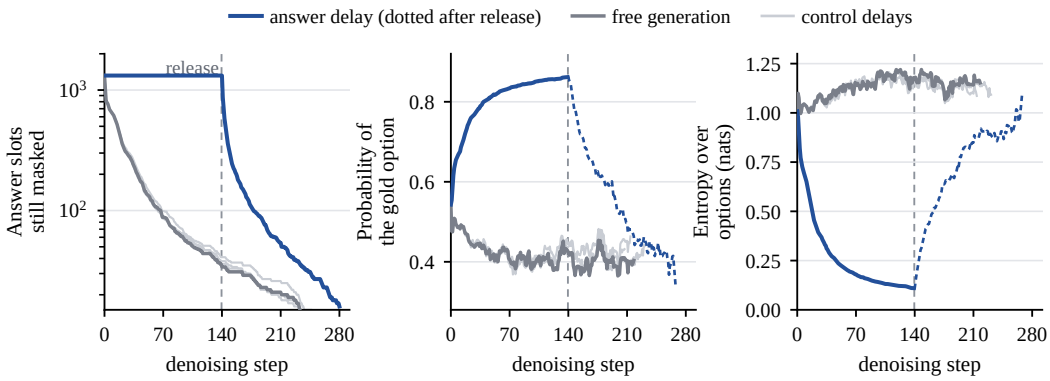

Figure 11: **What the answer slot predicts while it is masked** (terminal-choice scaffold, LLaDA-8B, 1,319 questions). Left: number of answer slots still masked. Middle and right: mean probability of the gold option and entropy over the four options, averaged over the slots still masked at each step (shown while at least 20 remain). Only the answer-delay arm before release averages all questions; the dotted continuation and the other arms describe the slots that have not yet become confident. The three control delays (timing-matched and two random-position controls) are indistinguishable from free generation.

Influence-matched control. The timing-matched control equates when a position is delayed but not how much the delayed token influences the rest of the response. A further control delays the reasoning position whose one-step influence on the still-masked predictions (mean Jensen–Shannon change after re-masking only that position) is closest to the answer's. On the 626 questions where this target is defined, answer delay exceeds the influence-matched delay by $+2.2$ pp (90% CI $[+0.8, +3.7]$) but does not separate from the timing-matched delay ($+0.5$ pp, 90% CI $[-0.8, +1.8]$), and the influence-matched delay itself lowers accuracy relative to the timing-matched delay (-1.8 pp, 90% CI $[-3.0, -0.6]$); it is an active intervention rather than a neutral placebo.

Tail-delay replay. The replay reuses the free-form machinery above with a different target. From the unrestricted reference (1,069 questions; the reference is token-identical to the run analysed in Section 3.2), the tail arm holds the last 32 positions of the reference response ineligible until step 448 of 512, and the span arm holds a contiguous span of 32 positions drawn, with a fixed seed per question, inside the central half of the reference response. Targets use the reference response length only; the 52 questions whose reference response is shorter than 80 tokens are inactive and take the reference outcome in both arms. A gated arm (w=16) was generated in the same job array. The analysis plan, with the primary contrast (tail minus span on questions that are end-anchored in the reference), the secondary family, and the interpretation rule, was fixed before generation. No held position was committed before release in either arm, the commitment budget was preserved on every question, and the held region was displaced on 62% of the end-anchored questions in the tail arm: the response ends before it, at the length of the gated solution, and the region is filled with terminators after release; on the other questions the response extends through the region after release, and the arms agree there as well. Table 10 reports all arms; Figure 12 shows one question. Within the secondary family (Holm adjusted), tail minus reference is $+31.4$ pp ([+23.2, +39.7]; p=5.7×10^{-11}), span minus reference $+0.5$ pp ([-2.1, +3.1]), gated minus reference $+32.0$ pp ([+23.7, +40.2]), and gated minus tail $+0.5$ pp ([-3.6, +4.6]) on the end-anchored questions; over all 1,069 questions, tail minus reference is $+5.8$ pp ([+4.0, +7.6]) and span minus reference -1.6 pp ([-2.7, -0.5]), a small cost of writing around a hole that also appears on the questions that are not end-anchored (-2.1 pp, [-3.3, -0.8]).

Table 10: Pre-specified tail-delay replay (LLaDA-8B, remaining 1,069 GSM8K questions, format-only, L=512, one job array). Rows for the reference-end-anchored questions also give the share of trajectories that are end-anchored under each arm, the median response length, and the median correlation between position and commitment order.

Questions		n	Unres.	Span held	Tail held	Gated
End-anchored in the reference	accuracy	194	32.0%	32.5%	63.4%	63.9%
	end-anchored		100%	100%	0%	0%
	median length		510	510	430	424
	position–order corr.		0.38	0.30	1.00	1.00
Not end-anchored	accuracy	875	86.9%	84.8%	87.0%	86.6%
All (ITT)	accuracy	1,069	76.9%	75.3%	82.7%	82.5%

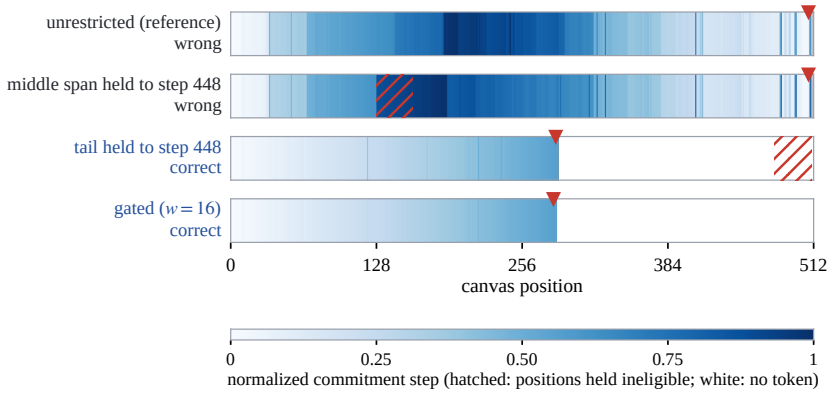

Figure 12: **One question under the four arms of the tail-delay replay.** The first question, in index order, that is end-anchored in the reference, wrong under the reference and the span arm, and correct under the tail arm and gating. Holding the middle span leaves the end fixed early and the middle filled in last; holding the tail makes the response left to right, and it ends before the held region.

Inducing the anchor. A third pre-specified experiment attempted the converse on 300 GSM8K questions that are not end-anchored in the reference: the canvas was truncated to the length of each question's gated solution by prefilling terminators, and the induce arm additionally prefilled the closing template at the end with the answer value cells left masked, while the control arm prefilled only the terminators. The primary contrast, induce minus control, is $+1.3$ pp (95% CI [-2.0, +4.7]; 14/10, p=0.541), so the prediction of a further loss from the planted template is not confirmed. The design does not isolate the template, because the control arm is already end-anchored on 95% of questions: with the end of the canvas known, the model fixes the closing sentence there first even when the room is exactly what its left-to-right solution used, and accuracy falls from 88.3% under the unrestricted reference (and gating) to 71.0% (-17.3 pp, [-22.7, -12.0]; exploratory). We report this as inconclusive with respect to the template and as a further instance of the loss that follows an early-fixed end.

Free-form replay. Without adding a slot, the unrestricted reference run defines for each question the mapped answer positions and the preceding reasoning positions. Two re-runs change eligibility only: the answer arm holds the answer positions ineligible until step 256 of 512, and the reasoning arm holds equally many reasoning positions chosen by minimum-cost matching of their reference commitment times. Targets use reference artifacts only; gold answers, correctness, and outcomes are never read. The active set, fixed before generation, contains the 498 of 1069 questions whose answer committed entirely before step 256 with at least 20 preceding tokens; for the others both arms take the reference outcome. On the active set the answer-minus-reasoning contrast is $+1.8$ pp (95% CI $[+0.0, +3.6]$; exact McNemar p=0.078), and over all questions it is $+0.8$ pp ($[+0.0, +1.7]$); both express the same nine net corrections. Both arms held every target until release and preserved the commitment budget, but the final answer moved away from the delayed positions in 8.8% of answer-

arm and 5.4% of reasoning-arm outputs, so the contrast estimates a reference-defined positional restriction. An unmodified replay reproduced the original trajectories on all 40 checked questions.

F ANSWER-FIRST MEASUREMENTS

Answer-first association on GSM8K. Among format-only unrestricted errors with a defined reasoning region, gating corrects 57.4% of answer-first and 23.0% of other errors (difference $+34.4$ pp, 95% CI $[+22.3, +46.5]$; Table 11); the net accuracy change differs by $+31.5$ pp (95% CI $[+21.7, +41.3]$). A logistic model of correction on answer-first status, gold-step tercile, log response length, and answer confidence gives an answer-first odds ratio of 1.62 (95% CI [0.82, 3.17]) and a standardized risk difference of $+9.3$ pp (95% CI [-4.2, +24.7]) against the crude $+34.4$ pp; adding answer-first status does not improve five-fold out-of-fold log loss (0.585 versus 0.585). Response length, which end-anchoring fixes at the canvas length, absorbs most of the association, consistent with Table 3.

Table 11: Gross corrections on 1,025 eligible format-only GSM8K responses, grouped by unrestricted trajectory. Error rescue divides corrections by baseline errors; rescue share divides them by all 86 corrections.

Group	n	Errors	Corrections	Error rate (%)	Rescue (%)	Share (%)
Answer-first	139	94	54	67.6	57.4	62.8
Other eligible	886	139	32	15.7	23.0	37.2

Timing definitions. Primary statistics use the complete-answer step t_{complete}, the first step at which the whole answer string is fixed; answer spans have a median length of 2 tokens on GSM8K and 2 on MATH. Under the median answer-token step t_{med}, responses can only move into the answer-first group: 24 of 1025 format-only GSM8K responses (2.3%) change status, and MATH answer-first rates rise by at most 6.5 percentage points. For the reasoning comparator, a minimum would instead ask whether the answer precedes every reasoning token and a maximum only whether it precedes the last; the median is less sensitive to isolated early or late tokens, but it is an ordering convention, not a semantic boundary of reasoning.

Answer-span mapping. Two annotators independently reviewed 183 stratified outputs (91 GSM8K, 92 MATH) without the model condition, correctness, or automatic label; their sheets agree on every span (Cohen's κ for mapping validity 1.00). The sample overrepresents fallback and boundary cases and informed the brace-balanced parser, so it is a development sample. The mapper reproduces 162/179 spans (90.5%), with a mean absolute t_{complete} error of 1.48 steps, and its answer-first label agrees with the human-span label wherever both are eligible (154/154). On MATH the mapper covers 9898 of 10386 item–condition outputs (95.3%; 82.2–100.0% across cells, lowest in long-canvas Dream conditions). Unmapped outputs are disproportionately canvas-filling and incorrect; assigning all of them to either category leaves every LLaDA policy contrast in answer-first rate negative. Accuracy is computed by a separate scorer on every item, and end-anchoring does not use the answer span, so mapping coverage affects only answer-first status and backfill.

G TERMINATION RULE

The main protocol retains EOS/EOT confidence suppression. A separate diagnostic on the first 250 GSM8K questions, generated on different hardware with an earlier sampler build, compares suppression on and off with the other settings fixed (Table 12). Without suppression, terminators are committed across the tail from the first steps on, the content is squeezed into the few positions that remain, and 90.8% of answers start within the first 20 positions, against 3.2% with suppression; accuracies are 52.8% and 75.6%. Both regimes fix the end of the canvas before the solution exists: with terminators when they can be selected, and with the closing sentence when they cannot. All primary results condition on suppression being on.

Table 12: GSM8K suppression diagnostic. Early start: the mapped answer begins before position 20; n_{order} counts outputs with a defined answer–reasoning order.

Setting	n	Accuracy	Early start	n_{order}	t_{complete} (t_{med})	τ_R	Answer-first
Suppression off	250	0.528	90.8%	22	0.352 (0.189)	0.504	68%
Suppression on	250	0.756	3.2%	242	0.490 (0.465)	0.274	13%