ERA-GAC FOR STABLE STRUCTURED REASONING WITH ATTENTION PRIORS AND GAIN-AWARE ENTROPY CONTROL

Anonymous authors
Paper under double-blind review

ABSTRACT

Large language models often fail on logical reasoning tasks not only by being wrong, but by being unstable, small shifts in training phase or attention dynamics can yield brittle internal inference that disproportionately harms structured question answering. We present ERA-GAC, a training-time method that stabilizes attention-based inference through (i) ERA, an additive log-prior over attention destinations that encodes length-aware structural constraints, and (ii) GAC, a gain-aware entropy/temperature controller that prevents late-phase collapse into overly sharp or overly diffuse attention regimes while keeping inference-time behavior fixed. On compute-matched training (540M parameters, 1.87B tokens), ERA-GAC yields statistically significant differences on six of nine tasks (five improvements, one regression) with paired tests and multiple-comparisons corrections. We find large gains on structured science QA benchmarks (SciQ, ARC-Easy, QASC) and smaller gains on commonsense/RC tasks, while the regression on BoolQ indicates when structural priors may interfere with passage-grounded entailment.

1 INTRODUCTION

Transformers (Vaswani et al., 2017) are the default backbone for sequence modeling, yet reasoning needs stable inference. Small late-phase changes often manifest as answer flips on structured multiple-choice QA. These flips are driven by brittle attention allocation (too sharp or too diffuse). We hypothesize that stabilizing attention as constrained inference improves robustness for structured MCQA.

We study architectural and training-time controls that improve logical and structured reasoning accuracy under fixed training compute, evaluating only compute-matched 540M / 1.87B tokens, greedy decoding, and a frozen harness. The evaluated suites largely instantiate deductive and abductive multiple-choice reasoning (ARC (Clark et al., 2018), SciQ (Welbl et al., 2017), QASC (Khot et al., 2020)) and passage-grounded entailment (BoolQ (Clark et al., 2019)).

Contributions

- **Stability-first logic reasoning:** We identify brittle attention regimes as a bottleneck for structured QA and formalize attention as constrained inference.
- **ERA-GAC:** A length-aware structural attention prior (ERA) combined with a gain-aware entropy controller (GAC) that stabilizes late-phase training.
- **Rigorous benchmarking:** Evaluation on structured reasoning suites with strictly matched compute, paired tests (McNemar, 1947), Holm correction (Holm, 1979), and explicit failure modes.

2 Background: Attention as Constrained Inference

Consider a single attention head with query $q_i \in \mathbb{R}^d$, keys $k_j \in \mathbb{R}^d$, values $v_j \in \mathbb{R}^d$, and logits $\ell_{ij} = \langle q_i, k_j \rangle / \sqrt{d}$. Standard attention (Vaswani et al., 2017) computes a categorical distribution

$$a_i = \mathrm{softmax}(\ell_i) \in \Delta^{n-1}, \quad o_i = \sum_{j=1}^{n} a_{ij} v_j. \tag{1}$$

We view attention as a KL-regularized MAP estimate:

Proposition 1 (Attention as KL-regularized MAP)**.** *Let $\pi_i \in \Delta^{n-1}$ be a prior over attention mass for query position i. For temperature $\tau > 0$, the solution to*

$$a_i^\star = \arg\max_{a \in \Delta^{n-1}} \langle a, \ell_i \rangle - \tau \, \mathrm{KL}(a \,\|\, \pi_i) \tag{2}$$

is $a_i^\star = \mathrm{softmax}((\ell_i + \tau \log \pi_i)/\tau)$ (up to an additive constant in logits).

This perspective maps attention to selecting evidence links, where the log-prior acts as a soft structural constraint. While not full symbolic logic, it provides a structured inference bias compatible with verifier hooks and constrained generation.

3 Method: ERA-GAC

3.1 Overview

High-level idea ERA-GAC augments a standard Transformer block (Vaswani et al., 2017) with two orthogonal mechanisms (**ERA** and **GAC**) that act sequentially on attention logits before the softmax operation.

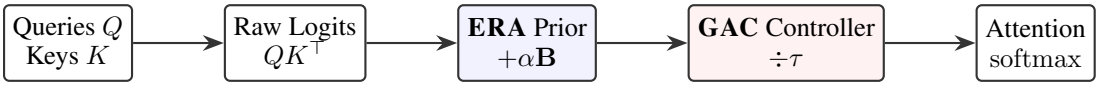

Figure 1: Conceptual overview of the ERA-GAC forward pass. The ERA prior provides structural constraints via an additive bias, while the GAC controller provides stability gating by dynamically scaling the logits.

The design is explicitly targeted at the "late-phase" training regime, where gradients are small and the model is prone to brittle shifts in attention allocation that disproportionately harm structured QA.

Implementation invariants To ensure comparability and exact recomputation, we enforce: (i) a single, frozen definition of the scoring rule for multiple-choice evaluation (primary: $-\sum$ token logprobs; secondary: mean logprob sensitivity), (ii) a single, frozen definition of the bias injection site (*additive* to attention logits), and (iii) full per-example logging such that headline accuracies can be recomputed exactly from logs.

Notation We use n for sequence length, R for the number of regimes, and B (overloaded) for the number of positional basis blocks. Vectors are row-vectors by default. For a matrix M, $M_{i:}$ denotes row i.

3.2 ERA (Entropic Regime Alignment): Structural log-priors

Goal We want a *length-aware* prior over key positions for each query token that captures block/segment structure and integrates into standard scaled dot-product attention as an additive logit bias.

Regime memberships We define R regimes that softly partition the sequence. For each position $t \in \{1, \dots, n\}$ we compute length-aware memberships $\mu_{t,r}$ based on learnable centers c_r:

$$\tilde{\mu}_{t,r} \;=\; \exp\Big(-\tfrac{1}{2\sigma^2} \, \|\phi(t;n) - c_r\|^2 \Big), \tag{3}$$

$$\mu_{t,r} \;=\; \frac{\tilde{\mu}_{t,r}}{\sum_{r'=1}^{R} \tilde{\mu}_{t,r'}}. \tag{4}$$

Entropic regime alignment (Sinkhorn) We align regimes to length-aware basis positional blocks $\psi_b(s;n)$ with an entropic optimal transport plan $T^\star \in \mathbb{R}_+^{R\times B}$. Given a distance cost matrix $C \in \mathbb{R}^{R\times B}$, we solve:

$$T^\star \;=\; \arg\min_{T\geq 0} \langle C, T\rangle + \varepsilon \sum_{r,b} T_{r,b}(\log T_{r,b} - 1)$$
$$\text{subject to } T\mathbf{1} = a,\; T^\top \mathbf{1} = b. \tag{5}$$

Constructing the prior and bias Given μ and $T^\star$, we define a prior distribution over keys for each query position t:

$$p_t(s) \;=\; \sum_{r=1}^{R} \mu_{t,r} \sum_{b=1}^{B} T^\star_{r,b} \, \psi_b(s;n). \tag{6}$$

We convert this to an additive logit bias by taking a stabilized log prior:

$$\mathcal{B}_{t,s} \;=\; \alpha \cdot \mathrm{RowZ}\big(\log(p_t(s) + \delta) \big), \tag{7}$$

where RowZ performs per-row z-scoring. Additive biases exploit the row-shift invariance of softmax, enabling robust normalization without affecting attention distributions. Z-scoring is preferable to a simple bounded bias or centering because it guarantees a consistent variance ($\sigma^2 = 1$) and dynamic range across all rows, preventing the log-prior from either vanishing or overwhelming the content logits regardless of the prior's initial entropy.

3.3 GAC: Gain-aware entropy control for stable inference

Motivation Even with a strong prior, late-phase training is fragile: small learning rates can cause attention to become (i) overly diffuse (entropy drifts upward), washing out token-specific structure; or (ii) overly sharp (entropy drifts downward), leading to brittle, high-variance gradients and memorization. GAC is a lightweight controller that adaptively modulates attention sharpness *during training* based on observed generalization gains, while keeping inference-time behavior fixed.

Design rationale We chose a gain-conditioned controller rather than fixed temperature scheduling for three reasons: (1) *Generalization-aware:* the controller only adjusts temperatures when validation metrics improve, avoiding premature sharpening that can cause overfitting; (2) *Stability:* projection to $[\tau_{\min}, \tau_{\max}]$ and EMA smoothing prevent runaway dynamics common in unconstrained temperature learning; (3) *Inference decoupling:* GAC operates only during training—temperatures are frozen at inference, eliminating any runtime overhead and ensuring reproducible evaluation.

Temperature parameterization For each layer ℓ (and optionally each head h), we introduce a scalar temperature $\tau^{(\ell)}$ that scales the attention logits:

$$A^{(\ell,h)}_{t,s} = \mathrm{softmax}_s\left(\frac{1}{\tau^{(\ell)}} \, \widetilde{L}^{(\ell,h)}_{t,s} \right). \tag{8}$$

When $\tau^{(\ell)} < 1$, attention sharpens (lower entropy); when $\tau^{(\ell)} > 1$, attention flattens. We constrain $\tau^{(\ell)}$ to a compact interval to avoid degeneracy: $\tau^{(\ell)} \in [\tau_{\min}, \tau_{\max}]$ with $0 < \tau_{\min} < 1 < \tau_{\max}$.

Gain signal Let m_k denote a held-out validation metric (e.g., negative cross-entropy or accuracy proxy) evaluated at checkpoints k. We define a smoothed "gain" signal using an exponential moving average (EMA):

$$\bar{m}_k = (1-\beta)\, \bar{m}_{k-1} + \beta \, m_k, \tag{9}$$

$$g_k = \bar{m}_k - \bar{m}_{k-1}. \tag{10}$$

Intuitively, g_k is positive when generalization is improving and negative when it is stalling or regressing.

Controller update rule GAC updates temperatures only when the gain warrants sharper attention. We use a projected deterministic multiplicative update (stable across scales):

$$\tau^{(\ell)} \leftarrow \Pi_{[\tau_{\min},\tau_{\max}]}\Big(\tau^{(\ell)} \cdot \exp(-\eta\, u_k\, \nabla_{\tau^{(\ell)}} \mathcal{L}_{\text{train}})\Big),$$
$$u_k = \mathbb{1}[g_k > \delta], \tag{11}$$

where η is a small controller step size, $\delta \geq 0$ is a dead-zone threshold, and Π is projection. This can be interpreted as "take a small sharpening step only when validation is still improving."

Entropy regularization To avoid runaway sharpening, we regularize attention entropy toward a target band. Let $H^{(\ell)}$ be the mean attention entropy in layer ℓ over a calibration batch. We add a penalty:

$$\mathcal{L}_{\text{ent}} = \lambda \sum_{\ell} \big[\, \max(0, H_{\min} - H^{(\ell)})^2 + \max(0, H^{(\ell)} - H_{\max})^2 \big]. \tag{12}$$

The full training objective is $\mathcal{L}_{\text{train}} + \mathcal{L}_{\text{ent}}$.

Inference-time behavior We do *not* run the controller at inference. At evaluation, $\tau^{(\ell)}$ is frozen to the value learned at the end of training.

3.4 Complexity and practical overhead

ERA adds $O(nR) + O(RB)$ to compute the prior; inference-time is a cached positional bias; no asymptotic change.

4 Benchmarks and evaluation for structured logical reasoning

We evaluate logical reasoning capabilities under a strict, compute-matched pre-specified protocol:

- **Models and Compute:** All compared language models (Transformer, ERA-GAC) share matched parameter counts (∼540M) and are trained for an identical budget ($T_{\text{match}} = 1.87\,\text{B}$ tokens) using the same optimizer and learning rate schedule.
- **Frozen Harness:** We use lm-eval v0.4.9.2 with identical task definitions and preprocessing.
- **Decoding and Scoring:** Decoding is strictly greedy (argmax, no sampling). Our primary scoring rule is the negative sum of token log-probabilities under teacher forcing.

These elements combine to isolate the value of structural regularization (Table 1 and Table 2).

4.1 Tasks and evaluation harness

Task families We group tasks into two pre-specified families: (F1) **Science/structured reasoning**: SciQ, ARC-Easy/Challenge, QASC, OpenBookQA; (F2) **Commonsense/RC**: HellaSwag, CommonsenseQA, BoolQ, WinoGrande. All tasks are evaluated zero-shot on full validation/dev splits, with exact split sizes documented in the logs.

Harness and decoding We use a frozen evaluation harness version (lm-eval v0.4.9.2) with identical task definitions and preprocessing for all models. Decoding is greedy (argmax) with no sampling.

Multiple-choice scoring (primary) For each choice c (rendered as a leading-space string), we score by negative sum of token log-probabilities under teacher forcing:

$$s_{\text{sum}}(c \mid x) = -\sum_{i=1}^{|c|} \log p_\theta(c_i \mid x, c_{<i}). \tag{13}$$

We predict $\arg\min_c s_{\text{sum}}(c \mid x)$. We also compute a length-normalized sensitivity score $s_{\text{mean}}(c \mid x) = \frac{1}{|c|} s_{\text{sum}}(c \mid x)$ and log both predictions. Headline numbers use s_{sum} unless noted.

Truncation All prompts are left-truncated to $(\texttt{max\_length} - \texttt{max\_gen\_toks})$ under the harness defaults. The per-example log records the truncation flag and effective prompt length.

4.2 Per-example logging and reproducibility

Logging and frozen harness We log per-example predictions sufficient to recompute all reported metrics exactly. We freeze the harness version (lm-eval v0.4.9.2) and scoring rules. Any deviation is flagged by a mismatch in prompt hashes and invalidates statistical tests.

4.3 Statistics

Paired tests For each task, we compare two models A and B using McNemar's test on paired correctness outcomes. Let n_{01} be the number of examples where A is correct and B is wrong, and n_{10} where A is wrong and B is correct. We compute the McNemar statistic with continuity correction:

$$\chi^2 = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}}, \tag{14}$$

and evaluate the associated p-value under a χ^2_1 distribution. We log the full contingency tables and bootstrap CIs, but report only the raw accuracy delta (Δ), the odds-ratio effect size ($\text{OR} = \frac{n_{01}}{n_{10}}$), and the Holm-corrected p-values in the main text for brevity.

Multiple comparisons Within each pre-specified family (F1, F2), we correct McNemar p-values using Holm–Bonferroni.

5 Results

This section reports the main scientific claim: under a fixed training token budget $T_{\text{match}} = 1.87$B tokens and matched parameter count ($\sim$540M parameters), ERA-GAC improves or matches zero-shot downstream accuracy relative to a Transformer baseline, operating entirely as structured inference without an explicit external solver.

5.1 Science Suite

Table 1 reports accuracies for the pre-specified task suite. For each task we additionally report raw accuracy delta (Δ), odds ratios, and Holm–Bonferroni corrected p-values (Table 2); we log contingency tables and bootstrap CIs.

Summary of findings Under strict compute matching (1.87B tokens), ERA-GAC achieves statistically significant differences (after Holm correction at $\alpha = 0.05$) on six of nine tasks (five improvements, one regression on BoolQ):

Strong gains on the Structured Science Reasoning Suite (F1): SciQ (+13.0pp, OR=3.26, rank 1), ARC-Easy (+10.9pp, OR=2.61, rank 2), and QASC (+4.1pp, OR=1.96, rank 5) all show large effect sizes with highly significant p-values. These tasks largely instantiate deductive and abductive multi-hop reasoning over science questions. OpenBookQA also falls in this suite.

Modest gains on commonsense/RC (F2): HellaSwag (+1.4pp, OR=1.46, rank 3) and CommonsenseQA (+3.3pp, OR=1.54, rank 6) show statistically detectable but smaller improvements.

Table 1: Compute-matched structured reasoning accuracy (540M, 1.87B tokens; greedy decoding). †Statistically significant difference after Holm correction (either direction) (α=0.05).

Task	ERA-GAC	Transformer	Δ (pp)
SciQ†	67.7	54.7	+13.0
ARC-Easy†	47.4	36.5	+10.9
QASC†	17.9	13.8	+4.1
ARC-Challenge	20.1	19.4	+0.7
OpenBookQA	15.2	15.0	+0.2
HellaSwag†	27.4	26.0	+1.4
CSQA†	23.3	20.0	+3.3
BoolQ†	58.6	61.2	**-2.6**
WinoGrande	49.7	49.3	+0.4
Macro avg.	35.3	32.9	+2.4

Table 2: Compute-matched paired McNemar test p-values (Holm-corrected within family); effect size via OR.

Task	Δ	OR	p_{Holm}
SciQ†	+13.0	3.26	$<10^{-15}$
ARC-Easy†	+10.9	2.61	2×10^{-7}
HellaSwag†	+1.4	1.46	5×10^{-7}
BoolQ†	-2.6	0.68	7×10^{-5}
QASC†	+4.1	1.96	0.0006
CSQA†	+3.3	1.54	0.004
WinoGrande	+0.4	1.04	0.80
ARC-Chall.	+0.7	1.11	0.87
OpenBookQA	+0.2	1.04	1.00

Passage-grounded entailment regression: BoolQ shows a -2.6pp regression (OR=0.68, rank 4, $p < 10^{-4}$). BoolQ requires yes/no classification of natural-language questions paired with passages. We hypothesize that structural priors can interfere with direct passage evidence integration when the structure is less compatible with regime alignment. Furthermore, truncation logging exists for these tasks, suggesting that interactions between passage length, truncation, and the ERA structural prior may exacerbate this regression.

Parity: WinoGrande (+0.4pp, $p = 0.804$), ARC-Challenge (+0.7pp, $p = 0.871$), and OpenBookQA (+0.2pp, $p = 1.0$) show no statistically detectable differences.

5.2 Discussion and Limitations

Scope Our strongest evidence is for compute-matched training and zero-shot evaluation at the chosen scale (1.87B tokens) and data mixture. Broader generalization to larger scales, other corpora, multilingual settings, and instruction-tuned regimes remains open.

Single-seed training The compute-matched comparison uses single-seed training runs due to resource constraints. While we report paired bootstrap CIs and McNemar tests with Holm correction to control statistical uncertainty in evaluation, we cannot quantify training-run variance. Multi-seed replication would strengthen causal claims.

Task-specific regressions While our findings are consistent with the hypothesis that ERA-GAC improves robustness on structured QA, we show a statistically significant regression on BoolQ (-2.6pp, OR=0.68, $p < 10^{-4}$), a passage-grounded QA task. We hypothesize that the regime-aligned prior impairs direct passage-question interaction when structural induction is less critical. We note that truncation logging exists for these tasks, and interactions between passage length, truncation, and the ERA structural prior may exacerbate this regression, though further analysis is needed.

Solver integration (future work) Because ERA-GAC controls inference sharpness, it presents a natural integration point for a symbolic verifier or formal solver constraint check: if a solver detects a logical contradiction, the model could selectively lower its temperature τ (tighten entropy band) and predictably re-decode deterministically. This tight loop aligns structural priors directly with explicit symbolic constraints.

6 DIAGNOSTICS

This section records the primary diagnostics we track to evaluate model stability and failure modes during training. We explicitly track three checks:

- **Regime collapse:** where $H(\mu_i)$ concentrates to near-zero for most tokens and the prior degenerates. We monitor Sinkhorn convergence statistics to detect this.
- **Prior domination:** where the bias magnitude overwhelms learned logits and attention ignores content. We track $\mathrm{std}(\mathbf{B})$ pre/post z-scoring.
- **Controller oscillation:** where τ swings erratically and harms stability. We track τ trajectories to detect bounds saturation.

7 RELATED WORK

Attention priors and positional structure Classic Transformers use absolute position encodings (Vaswani et al., 2017), while later methods add relative biases (Shaw et al., 2018), rotary embeddings (Su et al., 2021), or linear attention masks (ALiBi) (Press et al., 2021) to improve long-context evaluation. ERA is complementary: it constructs an additive logit bias normalized under softmax invariances to provide a structural log-prior.

Optimal transport and Sinkhorn normalization Entropic regularization enables fast approximate OT via Sinkhorn iterations (Cuturi, 2013). We use this as a doubly-stochastic normalization primitive to align regime mass to a positional basis without trivial collapse (Genevay et al., 2018).

Temperature scaling and controllers Temperature control is common for model calibration (Hinton et al., 2015) and smoothing representations. GAC is distinct in being explicitly gain-aware: it only sharpens attention when validation metrics improve, avoiding unprincipled drift and adapting to the model's precise learning trajectory.

Statistical testing in benchmark evaluation Given the brittleness of static MCQA comparisons, paired testing is essential. We rely on paired McNemar's tests (McNemar, 1947) supported by Holm's multiple-comparisons correction (Holm, 1979) and bootstrap CIs.

8 CONCLUSION

ERA-GAC shows that small architectural/training-time inductive biases, when paired with conservative, validation-conditioned control, can produce reliable gains under strict compute matching, but also surface task-specific regressions that would be easy to miss without paired testing. In our 1.87B-token, compute-matched setting, ERA-GAC shows statistically significant differences on six of nine pre-specified tasks (five improvements, one regression on BoolQ) after Holm–Bonferroni correction (Holm, 1979), with the largest benefits on structured science/knowledge tasks.

More broadly, we view this work as part of a shift in the field from chasing marginal benchmark deltas toward (i) explicitly stated inductive assumptions, (ii) evaluation protocols that quantify uncertainty and multiple comparisons, and (iii) mechanisms that trade expressivity for stability in the late phase. Progress on small and mid-scale models will increasingly depend on such "compute-accountable" interventions, and on reporting standards that make both improvements and regressions reproducible and interpretable.

REFERENCES

Christopher Clark, Kenton Lee, Ming-Wei Chang, Tom Kwiatkowski, Michael Collins, and Kristina Toutanova. Boolq: Exploring the surprising difficulty of natural yes/no questions. In *NAACL-HLT*, 2019.

Peter Clark, Isaac Cowhey, Oren Etzioni, Tushar Khot, Ashish Sabharwal, Carissa Schoenick, and Oyvind Tafjord. Think you have solved question answering? try arc, the ai2 reasoning challenge. In *arXiv preprint arXiv:1803.05457*, 2018.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, 2013.

Aude Genevay, Gabriel Peyré, and Marco Cuturi. Learning generative models with sinkhorn divergences. *International Conference on Artificial Intelligence and Statistics*, 2018.

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

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 1979.

Tushar Khot, Peter Clark, Michal Guerquin, Peter Jansen, and Ashish Sabharwal. Qasc: A dataset for question answering via sentence composition. In *AAAI*, 2020.

Quinn McNemar. Note on the sampling error of the difference between correlated proportions or percentages. In *Psychometrika*, 1947.

Ofir Press, Noah A. Smith, and Mike Lewis. Train short, test long: Attention with linear biases enables input length extrapolation. *arXiv preprint arXiv:2108.12409*, 2021.

Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. *arXiv preprint arXiv:1803.02155*, 2018.

Jianlin Su, Yu Lu, Shengfeng Pan, Bo Wen, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *arXiv preprint arXiv:2104.09864*, 2021.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, 2017.

Johannes Welbl, Nelson F. Liu, and Matt Gardner. Crowdsourcing multiple choice science questions. In *Workshop on Noisy User-generated Text*, 2017.

A Implementation Details

ERA Parameters and Formulas The normalized coordinate mapping is simply $\phi(t; n) = t/n$, mapping sequence indices to $[0, 1]$. We define basis blocks $\psi_b(s; n)$ as uniform non-overlapping indicators representing fractional chunks of the total context n, such that $\psi_b(s; n) = 1$ if $\frac{s}{n} \in [\frac{b-1}{B}, \frac{b}{B})$ and 0 otherwise. We set $R = 32$ semantic regimes and $B = 32$ basis positional blocks.

The distance cost matrix $C \in \mathbb{R}^{R\times B}$ for optimal transport is computed as the squared Euclidean distance between the learnable regime centers c_r and fixed basis centers $c_b = \frac{b-0.5}{B}$. Sinkhorn transport uses an entropic regularization hyperparameter $\varepsilon = 0.05$ and runs for 5 fixed iterations. The regime membership scale is $\sigma = 0.1$. The stabilized log prior uses $\alpha = 1.0$ and $\delta = 10^{-5}$ to prevent $\log(0)$ instability.

GAC Parameters The GAC EMA momentum is $\beta = 0.1$, the improvement threshold is $\delta = 0$, and the target entropy bands are $H_{\min} = 2.0, H_{\max} = 5.0$ (adjusted depending on baseline properties). The temperature is clipped between $\tau_{\min} = 0.5$ and $\tau_{\max} = 2.5$. The learning rate for the controller is $\eta = 0.01$.

Bias Injection Site The constructed ERA position bias $\mathcal{B}_{t,s}$ is additively injected into the pre-softmax attention logits computed by $QK^\top$ identically across all attention heads uniformly in all layers of the Transformer.

B Pseudocode

Algorithm 1 ERA Bias Construction

Input: Sequence length n, target size $R = 32, B = 32$
Parameters: Learnable centers $c \in \mathbb{R}^R$, $\sigma = 0.1, \varepsilon = 0.05, \delta = 10^{-5}, \alpha = 1.0$
Compute mappings $\phi(t; n) = t/n$ and blocks $\psi_b(s; n)$
$\mu_{t,r} \leftarrow \mathrm{softmax}_r(-\|\phi(t; n) - c_r\|^2/(2\sigma^2))$
Compute cost $C \in \mathbb{R}^{R\times B}$ as distances between c and basis centers c_b
$T^\star \leftarrow \mathrm{Sinkhorn}(C, \varepsilon, \text{iterations} = 5)$
$p_t(s) \leftarrow \sum_{r=1}^{R} \mu_{t,r} \sum_{b=1}^{B} T^\star_{r,b} \psi_b(s; n)$
$\mathcal{B}_{t,s} \leftarrow \alpha \cdot \mathrm{RowZScore}(\log(p_t(s) + \delta))$
Return: $\mathcal{B}$

Algorithm 2 GAC Controller Update

Input: Current temperatures τ, step k, gradients ∇_τ
State: Moving average validation metric $\bar{m}$
Hyperparameters: $\eta = 0.01$, $\delta = 0$, limits $[\tau_{\min}, \tau_{\max}]$, $\beta = 0.1$
if step k is a validation step **then**
    Evaluate validation metric m_k
    $g_k \leftarrow ((1 - \beta)\bar{m} + \beta m_k) - \bar{m}$
    $\bar{m} \leftarrow (1 - \beta)\bar{m} + \beta m_k$
    $u_k \leftarrow \mathbb{1}[g_k > \delta]$
end if
if $u_k = 1$ **then**
    $\tau \leftarrow \tau \cdot \exp(-\eta \cdot \nabla_\tau \mathcal{L}_{\text{train}})$
end if
$\tau \leftarrow \max(\tau_{\min}, \min(\tau_{\max}, \tau))$
At Inference: u_k is fixed to 0, τ is frozen, yielding zero overhead.

Table 3: Sample efficiency: ERA-GAC (1.87B tokens) vs Transformer (23.7B tokens). With 12.7× fewer training tokens, ERA-GAC matches or exceeds the longer-trained baseline on 5/9 tasks. This is **not** a compute-matched comparison.

Task	ERA-GAC (%)	Transformer (%)
SciQ	**67.7**	66.7
ARC-Easy	**47.4**	45.4
QASC	17.9	**19.8**
CSQA	23.3	**23.8**
BoolQ	58.6	**58.8**
HellaSwag	**27.4**	27.0
WinoGrande	**49.7**	48.1
OpenBookQA	**15.2**	13.4
ARC-Chall.	20.1	**21.1**

C Sample Efficiency (Secondary Evidence)

We include one non-matched comparison to illustrate sample-efficiency: ERA-GAC trained for 1.87B tokens vs a Transformer trained for 23.7B tokens (∼12.7× more tokens). **This is not used for our primary claim**

ERA-GAC matches or exceeds the longer-trained Transformer on 5/9 tasks (Table 3), with strongest advantages on SciQ (+1.0pp, trained on $\sim 1/12.7$ the data) and ARC-Easy (+2.0pp).

This suggests ERA+GAC can accelerate the emergence of structured knowledge behavior on early-stage evaluations. It does not imply superiority at equal compute or after extended training.