Reversal Invariance
A symmetry argument against the arrow of chain-of-thought

Mihir Sahasrabudhe
University of Illinois
mihirss2@illinois.edu

Abstract

We identify a fundamental symmetry in causal (autoregressive) language modeling (CLM): *reversal invariance*. Formally, the next-token prediction objective assigns the same loss to a corpus whether it is read forward or backward, showing that CLM pretraining is blind to sequence direction. This explains why models trained on reversed text can learn as effectively as forward-trained models, despite the inherent asymmetry of natural language. To address this limitation, we introduce *Asymmetry-Aware Pretraining (AAP)* as a conceptual framework, which augments the objective with penalties that favor forward-flowing text. We hypothesize that, if validated, AAP could better align representations with causal and temporal structure, while preserving standard language modeling performance. This perspective reframes pretraining from optimizing symmetric likelihood to explicitly capturing the arrow of language and reasoning.

1 Motivation

Reasoning is inherently asymmetric: premises lead to conclusions along an arrow that is not freely reversible. Human language reflects this asymmetry as well: sounds unfold in time, morphology encodes direction-sensitive dependencies, and discourse progresses from past to future. A training objective that ignores this arrow may overlook information essential for causal and temporal understanding.

Yet the standard **causal (autoregressive) language modeling (CLM)** objective [17, 2] is blind to direction. While architectures impose a causal mask at inference to restrict predictions to past tokens, this masking convention presupposes that text is read forward. At the level of *pretraining*, the next-token negative log-likelihood (NLL) loss itself does not distinguish forward text from its reversal. This raises a central question: if the objective is symmetric, to what extent can models internalize the irreversibility of language without additional training signals?

We argue, as a theoretical position, that if natural language exhibits time-asymmetric structure (phonotactics, morphology, agreement, discourse flow), then a direction-blind objective may systematically overlook it. This motivates our proposal of *asymmetry-aware objectives* as a conceptual direction.

2 Introduction

Large language models (LLMs) achieve strong results across diverse tasks, and *chain-of-thought (CoT) prompting* often improves performance by allocating more inference-time computation [25]. However, such inference-time gains do not alter the fundamental pretraining objective. We argue that the CLM loss is formally *reversal-invariant*: the next-token NLL is unchanged if a corpus is read backward and the vocabulary permuted accordingly. This invariance helps account for why models trained on reversed corpora can match forward-trained models in perplexity and learning dynamics [30].

Prior approaches have attempted to work around this limitation. Right-to-left (R2L) training performs comparably to left-to-right (L2R) in both MT and LM settings [21, 31]. Bidirectional models such as ELMo [15] and BERT [5] rely on independent or masked objectives, while XLNet [29] generalizes autoregressive pretraining through permutations. These methods either rely on symmetry or sidestep it, rather than addressing the absence of directionality in the base objective. Post-training methods such as reinforcement learning from human feedback (RLHF) [3, 13] inject asymmetry through preference signals, but only after pretraining is complete.

We introduce **Asymmetry-Aware Pretraining (AAP)** as a conceptual framework: a modification of the CLM loss that compares each forward se-

quence with its reversal and explicitly rewards the model for preferring the natural direction. While computationally more demanding, we suggest that AAP provides a principled direction for encoding irreversibility into pretraining itself, rather than relying solely on inference heuristics or post-hoc alignment. We hypothesize that, if validated, AAP could yield measurable gains on direction-sensitive benchmarks such as temporal commonsense, narrative ordering, and causal question answering, while preserving general LM performance. In doing so, we aim to reframe pretraining from optimizing symmetric likelihood to explicitly capturing the *arrow of language and reasoning*.

3 Related Work

3.1 Autoregressive training and reversal

Autoregressive (AR) language models minimize next-token negative log-likelihood (NLL) over tokenized text sequences [17, 2]. While AR architectures apply a *causal mask* at inference to restrict attention to past tokens, this convention presupposes a forward reading order and does not by itself endow the *training objective* with directionality [19, 33].[1] Prior engineering results report that "right-to-left" (R2L) or reversed training performs comparably to left-to-right (L2R) in both MT and LM settings, and recent studies show that training from scratch on fully reversed corpora can match forward learning curves ("reverse modeling") [30]. To our knowledge, a clean, general statement of *reversal invariance* for AR objectives—explicitly accounting for tokenizer stability, vocabulary permutations, and positional encodings—has not been clearly formalized; here we aim to articulate this gap.

3.2 Reversal Curse and directional generalization

The *Reversal Curse* documents brittle, direction-sensitive generalization—e.g., models trained on "A is B" may fail to answer "B is A" [1, 6, 7]. Follow-up analyses study gradient dynamics and architectural factors behind such asymmetries and propose data or training adjustments. Our perspective is complementary: we trace part of this brittleness to a symmetry of the AR objective itself and hypothesize adding explicit directional signals at *pretraining* time.

3.3 Chain-of-thought as compute at inference

Chain-of-thought (CoT) prompting improves accuracy by allocating additional inference-time computation and exposing intermediate structure [26, 10]. We do not treat CoT as evidence that the base AR objective encodes causality; rather, we view it as orthogonal to training-time directionality. Our proposals aim to make *representations* more direction-aware during pretraining, potentially complementing CoT with stronger directional priors.

3.4 Information-theoretic lens on directionality

For stationary sources, Shannon entropy rate—and thus the ideal perplexity floor—is invariant under time reversal [4, 11]. Natural language, however, exhibits asymmetric constraints (phonotactics, morphology, syntax, discourse) that can be captured by a *time-reversal divergence*, i.e., the per-token Kullback–Leibler divergence between forward and reversed path measures. Because AR NLL optimizes a symmetric target, it is insensitive to this divergence, motivating our proposal of asymmetry-aware penalties as a conceptual direction.

3.5 Preference optimization and post-training

Preference-optimization methods such as RLHF introduce asymmetry *post-training* by rewarding response formats and styles aligned with human preferences [3, 13, 16]. We hypothesize that providing *explicit directional signals during pretraining* could, if validated, reduce reliance on post-hoc preference data for direction-sensitive behaviors, while remaining compatible with downstream preference optimization [20].

3.6 Attention Masking and Directionality

Causal masking in Transformer decoders ensures tokens attend only to their predecessors, instilling directional structure at inference time. Foundational work [24] outlines this architecture, while more recent analyses probe its deeper effects. For example, Wu et al. [27] demonstrate that attention masks interact with LayerNorm to shape representation dynamics and mitigate rank collapse. Further, self-attention has been interpreted through a causal lens, suggesting that attention patterns can implicitly encode structural dependencies [18]. Together, these findings highlight how architectural

[1] We use "causal" in the standard architectural sense (masking future tokens) rather than to suggest an intrinsic, irreversible arrow in the NLL objective.

constraints establish directionality, yet still do not impose it during pretraining.

3.7 Symmetry and Invariance in Deep Learning

Theoretical work on symmetry in neural networks has shown that parameter-space symmetries impact learning dynamics, generalization, and optimization landscapes [34]. Methods for enforcing, discovering, and breaking symmetries have been proposed across model classes [12]. However, symmetric objectives themselves may fail to respect natural asymmetries, underscoring our argument that explicit directional signals may be needed in language modeling.

3.8 Connections to Causal Inference

Bridging causal inference with Transformer models, Zhang et al. [32] uncover a formal duality between attention mechanisms and structural causal models, enabling zero-shot causal discovery. Surveys in causal deep learning elucidate how causal inference frameworks can inform and improve generalization in neural systems [8]. These perspectives reinforce our position that capturing real-world causal and temporal asymmetries may require going beyond symmetric prediction objectives.

4 Formalizing Reversal

4.1 Preliminaries

Let Σ be a finite alphabet and Σ^* the free monoid of finite strings under concatenation. A (training) corpus is a finite multiset $D \subset \Sigma^*$. A tokenizer is a map $\tau : \Sigma^* \to V^*$ from strings to token sequences over a finite vocabulary V. For $s \in \Sigma^*$ write $\tau(s) = z = (z_0, \dots, z_{m-1}) \in V^*$.

An autoregressive (AR) model with parameters θ defines conditional distributions $p_\theta(z_{k+1} \mid z_{\le k})$ with joint factorization

$$\log P_\theta(z) = \sum_{k=0}^{m-2} \log p_\theta\big(z_{k+1} \,\big|\, z_{\le k}\big). \quad (1)$$

The negative log-likelihood (NLL) risk on corpus D is

$$\mathcal{L}_{\mathrm{NLL}}(\theta; \tau, D) = \mathbb{E}_{s\sim D}\left[-\log P_\theta\big(\tau(s)\big)\right]. \quad (2)$$

4.2 Reversal on strings and tokens

Definition 4.1 (String reversal)**.** The reversal operator $T : \Sigma^* \to \Sigma^*$ is defined by

$$T(x_0 x_1 \cdots x_{m-1}) = x_{m-1} \cdots x_1 x_0,$$

extended multiplicatively as $T(xy) = T(y)T(x)$. It is an involution: $T \circ T = \mathrm{id}$.

Given a tokenizer τ trained on D, let τ_T denote a tokenizer trained on $T(D)$. Define the token-sequence reversal $\mathrm{rev} : V^* \to V^*$ by

$$\mathrm{rev}(z_0, \dots, z_{m-1}) = (z_{m-1}, \dots, z_0).$$

Definition 4.2 (Tokenization stability under reversal)**.** We say τ and τ_T are *stable under reversal* if there exists a vocabulary bijection $\pi : V \to V_T$ such that, for all $s \in D$,

$$\tau_T\big(T(s)\big) = \pi\big(\mathrm{rev}(\tau(s))\big). \quad (3)$$

Discussion. For frequency-based tokenizers (e.g., BPE), full-string reversal preserves adjacent symbol statistics up to order. In large-data regimes, merge trees concentrate so that τ and τ_T differ only by a near-permutation.[2]

4.3 Permutation equivariance of the model

Let $E \in \mathbb{R}^{|V|\times d}$ be the input embedding matrix and $W \in \mathbb{R}^{|V|\times d}$ the (tied or untied) output projection. For a vocabulary permutation $\pi : V \to V$ with permutation matrix $P_\pi \in \{0,1\}^{|V|\times|V|}$, define the parameter reindexing

$$\Phi_\pi(\theta): \quad E' = P_\pi E, \qquad W' = W P_\pi^\top, \quad (4)$$

leaving all other weights unchanged (self-attention and MLP blocks are token-id agnostic).

Lemma 4.3 (Vocabulary permutation equivariance)**.** *For any sequence $z \in V^*$ and prefix $c \in V^*$, we have*

$$p_\theta(z \mid c) = p_{\Phi_\pi(\theta)}\big(\pi(z) \mid \pi(c)\big), \quad (5)$$

$$\log P_\theta(z) = \log P_{\Phi_\pi(\theta)}\big(\pi(z)\big). \quad (6)$$

Proof sketch. Permuting rows of E relabels input token embeddings; permuting columns (or rows, if tied) of W relabels output logits. All intermediate computations are invariant to token identities, so conditionals are preserved under consistent relabeling. □

4.4 Handling positions

Let $J_m \in \{0,1\}^{m\times m}$ be the *index-reversal* matrix $(J_m)_{i,j} = \mathbf{1}\{i + j = m - 1\}$, which maps position j to $m - 1 - j$.

[2] Tie-breaking and finite-data effects introduce small deviations; our results apply exactly under (3) and approximately otherwise.

Remark 4.4 (Positional encodings). With relative or rotary positional encodings, reversing the token order and the causal mask is functionally equivalent to a forward pass on the reversed sequence; no parameter change is needed. With absolute learned position embeddings $P \in \mathbb{R}^{m_{\max} \times d}$, one can compose a fixed index-flip $P'(j) = P(m_{\max} - 1 - j)$, or treat the flip as part of the data pipeline. Our invariance result assumes either relative/rotary encodings or that an index-flip is available.

4.5 Reversal invariance of the AR objective

Theorem 4.5 (Reversal invariance up to reparameterization)**.** *Assume tokenization stability* (3) *and a positional encoding scheme satisfying Remark 4.4. Then, as a theoretical observation, there exists a parameter map* Ψ *(comprised of* Φ_π *and, if needed, a fixed position index flip) such that*

$$\mathcal{L}_{\mathrm{NLL}}(\theta; \tau, D) = \mathcal{L}_{\mathrm{NLL}}\big(\Psi(\theta); \tau_T, T(D)\big). \quad (7)$$

Proof sketch. Fix $s \in D$ and write $z = \tau(s)$ and $z^T = \tau_T(T(s))$. By (3), $z^T = \pi(\mathrm{rev}(z))$. The AR log-likelihood (1) on z is a sum of local conditionals over adjacent prefixes. Running the same network on the reversed token sequence with reversed causal mask yields the same chain of conditionals, up to (i) relabeling tokens by π and (ii) a fixed index flip for absolute positional embeddings. By Lemma 4.3, there exists $\Phi_\pi(\theta)$ so that $\log P_\theta(z) = \log P_{\Phi_\pi(\theta)}(z^T)$. Averaging over $s \sim D$ yields (7). □

Corollary 4.6 (Loss landscapes and minima)**.** *Under the hypotheses of Theorem 4.5, empirical risk landscapes on* (τ, D) *and* $(\tau_T, T(D))$ *are identical up to the smooth reparameterization* Ψ*. In particular, sets of minimizers are in bijection via* Ψ*, and matched training runs are* predicted *to exhibit indistinguishable learning curves when evaluated on their respective domains.*

Remark 4.7 (Scope and approximation). Exact equality (7) holds under (3) and position handling as in Remark 4.4. In practice, the equality is approximate due to tokenizer tie-breaks, finite-sample effects, and implementation details (e.g., special tokens, padding). These deviations are typically small in large-corpus regimes, consistent with empirical reports of near-identical forward/reverse learning curves.

Interpretation. Equation (7) shows that next-token NLL is *blind to direction*: replacing D by its reversal $T(D)$ and retokenizing yields the same training problem up to a relabeling of tokens (and a fixed positional index flip). Thus any apparent left-to-right "arrow of inference" in chain-of-thought is not explained by the base objective alone; it likely arises from properties of the data distribution or from additional, explicitly asymmetric training signals.

5 An Information-Theoretic Perspective

5.1 Entropy rate and perplexity

Let $\{X_t\}_{t \in \mathbb{Z}}$ be the stationary stochastic process induced by drawing token sequences from the data distribution. The *entropy rate* of this process is

$$\begin{aligned} h &= \lim_{n \to \infty} \frac{1}{n} H(X_1^n), \\ H(X_1^n) &= -\sum_{x_1^n} P(x_1^n) \log P(x_1^n). \end{aligned} \quad (8)$$

where $X_1^n = (X_1, \ldots, X_n)$.

In practice, the entropy rate sets the information-theoretic lower bound for next-token prediction. The *perplexity* of a model is

$$\mathrm{PPL}(p_\theta) \;=\; \exp\left(\lim_{n \to \infty} \frac{1}{n} \mathbb{E}\big[-\log p_\theta(X_1^n) \big] \right). \quad (9)$$

For a perfect model, $\mathrm{PPL} = \exp(h)$. Thus entropy rate h is the statistical ceiling: no model trained purely on likelihood can beat this bound.

5.2 Reversal symmetry of entropy rate

Let $\{X_t^R\}$ denote the time-reversed process, with

$$P^R(x_1^n) = P(x_n, \ldots, x_1).$$

Because Shannon entropy is invariant under symbol reordering inside the probability distribution, one obtains

$$h(X) \;=\; h(X^R). \quad (10)$$

In other words, the entropy rate — and therefore the limiting perplexity — is the same for forward and reversed processes. This supports the intuition that autoregressive training alone may not distinguish forward from backward text.

5.3 Time-reversal divergence: measuring asymmetry

Despite identical entropy rates, natural language is not statistically reversible. A rigorous measure

of this asymmetry is the *time-reversal divergence* (sometimes called entropy production rate):

$$\mathcal{A} = \lim_{n\to\infty} \frac{1}{n} D_{\mathrm{KL}}\big(P(X_{1:n}) \,\|\, P^R(X_{1:n})\big), \quad (11)$$

where D_{KL} is the Kullback–Leibler divergence.

Remark 5.1. If $\mathcal{A} = 0$, the process is statistically indistinguishable from its reversal (full symmetry). If $\mathcal{A} > 0$, there is a detectable arrow of time in the data distribution. For natural languages, prior linguistic evidence suggests that $\mathcal{A}$ is nonzero: e.g., phonotactic rules, syntactic dependencies, and discourse structures are direction-sensitive.

5.4 Implication

Equations above suggest:

- Entropy rate h and perplexity are *reversal-invariant*.
- But the KL-based asymmetry $\mathcal{A}$ is strictly positive for human language.

Thus we argue that the autoregressive NLL objective optimizes toward a quantity that is effectively *blind to direction* (h), while overlooking the genuine asymmetry present in data ($\mathcal{A}$). This gap motivates our proposal to consider explicitly encoding the arrow of time into model objectives. [14]

6 Baking the Arrow of Time into the Objective

6.1 Motivation

Section 5 argued that the autoregressive NLL objective is invariant under sequence reversal: it optimizes toward the entropy rate h, which is symmetric, while overlooking the time-reversal divergence $\mathcal{A}$, which captures genuine asymmetry in natural language. To move beyond direction-blind pattern matching, we suggest considering training objectives that explicitly encode irreversibility.

6.2 General formulation

Let $\mathcal{L}_{\mathrm{NLL}}(\theta)$ denote the standard loss. We outline a conceptual augmentation with an asymmetry-aware regularizer $\mathcal{A}_\star(\theta)$:

$$\mathcal{J}(\theta) = \mathcal{L}_{\mathrm{NLL}}(\theta) + \lambda\, \mathcal{A}_\star(\theta), \quad (12)$$

where $\lambda > 0$ controls the tradeoff. The choice of $\mathcal{A}_\star$ determines how the arrow of time might be enforced. [23, 9]

6.3 Candidate regularizers

Likelihood-ratio margin. For a sequence x_1^n, compare forward and reverse likelihoods:

$$\Delta_{\mathrm{LLR}}(x_1^n) = \log P_\theta(x_1^n) - \log P_\theta(T(x_1^n)). \quad (13)$$

One could require $\Delta_{\mathrm{LLR}} > \delta$ on average, penalizing models that assign comparable probability to forward and reversed text.

Conditional-entropy gap. Define the empirical forward and backward conditional entropies:

$$\widehat{H}_f = \mathbb{E}\big[-\log p_\theta(x_{t+1} \mid x_{\le t})\big], \quad (14)$$

$$\widehat{H}_b = \mathbb{E}\big[-\log p_\theta(x_t \mid x_{\ge t+1})\big]. \quad (15)$$

Linguistic evidence suggests $\widehat{H}_f < \widehat{H}_b$. One could penalize models when this inequality is violated, encouraging that predicting the future from the past is consistently easier than the reverse.

Monotone latent potential. Let $\phi(h_t)$ be a scalar probe of the hidden state at position t. One possible constraint is $\phi(h_{t+1}) \ge \phi(h_t)$, encouraging that the model's internal notion of "progress" grows monotonically as it processes a sentence in the forward direction.

6.4 Discussion

Each of these regularizers conceptually incorporates the *arrow of time* into training, biasing the model toward forward-directed structure. Unlike RLHF, which rewards stylistic formats of reasoning, these methods would aim to encode irreversibility at the level of the learning objective itself. We hypothesize that such asymmetry-aware training could help models move beyond the ceiling of direction-blind statistical pattern matching, toward representations that more faithfully capture causal, irreversible reasoning.

6.5 Toy Example: The Q→U Rule

A simple illustration of asymmetry in natural language comes from English spelling. In nearly all words, the letter *Q* is followed by the letter *U*. This creates a strong directional dependency:

- Forward prediction: given context ending in "Q", the model can predict "U" with probability ≈ 1.
- Reverse prediction: given context starting with "U", there is no strong guarantee that the preceding token was "Q"; "U" can follow many other letters.

Formally, let X_t denote characters in text. The forward conditional entropy

$$H(X_{t+1} \mid X_t = \text{“Q”}) \approx 0,$$

while the backward conditional entropy

$$H(X_t \mid X_{t+1} = \text{“U”}) > 0,$$

since many predecessors of “U” are possible. Thus $H_f < H_b$ for this process, exhibiting a nonzero time-reversal divergence $\mathcal{A} > 0$.

Implication. The Q→U rule demonstrates concretely why the arrow of time matters: the asymmetry is not incidental formatting but an intrinsic property of the data distribution. Conceptually encoding this irreversibility into training objectives (via entropy-gap penalties or likelihood-ratio margins) could provide a way for models to capture such structural asymmetries rather than treating forward and backward text as equivalent.

7 Asymmetry-Aware Pretraining

The reversal invariance of next-token training shows that autoregressive objectives are *direction-blind*. To capture the intrinsic asymmetry of natural language, we introduce *Asymmetry-Aware Pretraining* (AAP) as a conceptual framework: augmenting the standard NLL loss with a small regularization term that privileges forward sequences over their hypothetical reversals. We present a theoretical formalization of this idea below.

7.1 Forward preference margin

Definition 7.1 (Forward likelihood margin)**.** Let P be the data distribution over token sequences $x \in \mathcal{X}$, and let $T : \mathcal{X} \to \mathcal{X}$ denote sequence reversal. For an AR model p_θ, define the *forward likelihood margin*:

$$\Delta_\theta(x) \;=\; \log p_\theta(x) \;-\; \log p_\theta(Tx).$$

Problem 1 (Conceptual constrained maximum likelihood). Minimize forward NLL

$$\mathcal{L}_{\mathrm{NLL}}(\theta) \;=\; \mathbb{E}_{x\sim P}\big[-\log p_\theta(x)\big]$$

subject to a forward preference constraint

$$\mathbb{E}_{x\sim P}\big[\Delta_\theta(x)\big] \;\geq\; \delta,$$

for some margin $\delta > 0$.

7.2 AAP objective

Definition 7.2 (AAP penalized objective)**.** For $\lambda \geq 0$, define the augmented objective

$$\mathcal{J}_\lambda(\theta) \;=\; \mathcal{L}_{\mathrm{NLL}}(\theta) + \lambda\,\mathbb{E}_{x\sim P}\big[\,\max\big(0,\;\delta - \Delta_\theta(x)\big)\,\big]. \tag{16}$$

Theorem 7.3 (Conceptual properties and bounded tradeoff)**.** *Assume feasibility of Problem 1. Then, in theory:*

(i) There exists $\lambda^\star \geq 0$ such that global minimizers of $\mathcal{J}_{\lambda^\star}$ coincide with the constrained optima.

(ii) Let

$$\theta^0 := \arg\min_\theta \;\mathcal{L}_{\mathrm{NLL}}(\theta), \tag{17}$$

$$\theta_\lambda := \arg\min_\theta \;\mathcal{J}_\lambda(\theta). \tag{18}$$

Then

$$\begin{aligned}&\mathcal{L}_{\mathrm{NLL}}(\theta_\lambda) - \mathcal{L}_{\mathrm{NLL}}(\theta^0)\\ &\leq\; \lambda\,\mathbb{E}_{x\sim P}\big[\max\big(0,\;\delta - \Delta_{\theta^0}(x)\big)\big]\,.\end{aligned} \tag{19}$$

Hence, in principle, perplexity inflation is bounded by λ and the slack of the baseline.

7.3 Equivalent cross-entropy view

Proposition 1 (Cross-entropy gap identity). Define the reversed model $p_\theta^R(x) := p_\theta(Tx)$. Then

$$\mathbb{E}_{x\sim P}[\Delta_\theta(x)] \;=\; H(P, p_\theta^R) - H(P, p_\theta),$$

where $H(P, Q)$ is cross-entropy. Thus, in theory, the forward preference constraint can be viewed as requiring that p_θ fits the data strictly better than its reversal by margin δ.

7.4 Regularizer variants

Definition 7.4 (AAP regularizers)**.** Several possible conceptual instantiations of asymmetry-aware penalties can be considered:

- **Likelihood-margin hinge:**

$$\mathcal{A}_{\mathrm{LLR}}(\theta) \;=\; \mathbb{E}_{x\sim P}\big[\max\big(0,\;\delta - \Delta_\theta(x)\big)\big]\,.$$

- **Conditional-entropy gap:** intended to encourage forward prediction to be easier than backward:

$$\mathcal{A}_{\Delta H}(\theta) \;=\; \max\big(0,\;\gamma - (\widehat{H}_b(\theta) - \widehat{H}_f(\theta))\big),$$

where $\widehat{H}_f$ and $\widehat{H}_b$ are empirical forward/backward conditional entropies.

- **Monotone progress:** for a probe ϕ on hidden states h_t,

$$\mathcal{A}_{\text{mono}}(\theta) \;=\; \mathbb{E}\Big[\sum_t \max(0,\; \phi(h_t) - \phi(h_{t+1}))\Big]$$

7.5 Predictions

Corollary 7.5 (Hypothesized signatures of AAP)**.** *Minimizers of $\mathcal{J}_\lambda$ with the above regularizers are* intended*, in theory, to exhibit:*

1. *Positive forward likelihood margin $\mathbb{E}[\Delta_\theta] \geq \delta$, biasing the model toward forward sequences.*
2. *Enlarged forward/backward entropy gap $\widehat{H}_b - \widehat{H}_f \geq \gamma$.*
3. *Hidden-state probes that reflect monotone forward progress.*

We hypothesize these signatures could correlate with improved accuracy on direction-sensitive reasoning tasks (temporal commonsense, causal QA, narrative ordering) at comparable perplexity.

7.6 Discussion and Future Directions

The Asymmetry-Aware Pretraining (AAP) framework is presented here as a conceptual approach: introducing an explicit penalty term that could bias autoregressive models toward preferring the natural forward direction of sequences. While the theoretical formulation is straightforward, several open questions remain.

Computational considerations. AAP requires evaluating both the forward sequence x and its reversal Tx in order to compute the likelihood margin $\Delta_\theta(x) = \log p_\theta(x) - \log p_\theta(Tx)$. This effectively doubles the cost of each batch relative to standard next-token prediction. In large-scale pretraining, such overhead would be nontrivial. The hyperparameter λ in the penalized objective $\mathcal{J}_\lambda(\theta)$ therefore controls not only the trade-off between NLL and the asymmetry term, but also the practical return on this additional computation.

Expected benefits. We do not expect AAP to substantially improve broad metrics such as perplexity. Current LMs already absorb a strong implicit forward bias from training on overwhelmingly forward-directed text, and $\Delta_\theta(x)$ is likely positive for most examples. Thus the penalty $\max(0, \delta - \Delta_\theta(x))$ may often be inactive, yielding a sparse signal. Instead, we hypothesize that, if validated, AAP could transform this implicit bias into an *explicit*, robust representational feature. Standard NLL encourages local coherence, but is agnostic to directionality; AAP is intended to encourage alignment with causal and temporal structure.

Future empirical validation. The central claims of this work remain hypotheses. Future work might test AAP in controlled experiments, comparing baseline and AAP-trained models on direction-sensitive benchmarks. Promising candidates include:

- **Causal QA**: datasets such as COPA that require distinguishing cause from effect.
- **Temporal and narrative ordering**: e.g., the Story Cloze Test or event sequencing tasks.
- **Commonsense physical reasoning**: benchmarks that probe knowledge of irreversible processes and event dynamics.

Outlook. The key open question is whether the computational overhead of AAP would be justified by measurable gains on these tasks. If validated, asymmetry-aware objectives could represent a principled step toward aligning language model training with the irreversible arrow of human reasoning. [28, 22]

8 Discussion and Implications

We argue that autoregressive next-token prediction is *reversal-invariant*, and therefore blind to the directional asymmetries that characterize natural language. This theoretical perspective suggests why chain-of-thought (CoT) prompting can appear directional while ultimately emerging from a symmetric pretraining objective. Empirical findings such as the Reversal Curse illustrate the potential practical impact: models trained under symmetric objectives often struggle to generalize logically equivalent statements across directions.

We outline three hypothesized implications:

1. **Interpretation of CoT.** Performance gains under CoT may be better understood as evidence of improved inference-time search and recall, rather than proof that the model has internalized a genuine arrow of reasoning. Our framework situates CoT as compute-at-inference layered on top of direction-blind pretraining.
2. **Scaling and ceilings.** Larger models and more data alone are unlikely to break reversal

invariance. Without objectives that explicitly encode irreversibility, pretraining may converge to representations that are statistically powerful but direction-agnostic, suggesting a ceiling on tasks that require causal or temporal grounding.

3. **Asymmetry-aware objectives.** By introducing directional penalties—likelihood-ratio margins, conditional-entropy gaps, or monotone latent potentials—models could be biased toward forward directionality. We hypothesize this would produce measurable signatures (e.g., larger forward-vs-reverse likelihood gaps) and possibly improved performance on direction-sensitive benchmarks such as temporal commonsense reasoning and causal question answering.

Taken together, these hypotheses point toward a shift in emphasis: from pursuing incremental gains on symmetric objectives, to designing pretraining signals that explicitly capture the inherent asymmetry of reasoning.

9 Conclusion

We have argued for a formal treatment of *reversal invariance* in autoregressive language modeling, showing that the next-token objective is invariant to string reversal up to vocabulary permutation and positional reindexing. From an information-theoretic perspective, this helps explain why entropy rates and perplexity floors are symmetric, even though natural language exhibits nonzero time-reversal divergence. To address this gap, we introduced *Asymmetry-Aware Pretraining (AAP)* as a conceptual framework: augmenting the NLL objective with penalties that privilege forward order.

Our contribution is deliberately theoretical: we have not implemented AAP at scale, nor do we claim empirical validation. Rather, our goal has been to formalize the symmetry, articulate its implications, and outline hypotheses for future testing. By framing this perspective, we aim to make the conceptual problem visible to the community and to encourage empirical follow-ups that evaluate the feasibility and impact of asymmetry-aware training.

10 Limitations

This work is a **theoretical position paper**. We did not conduct empirical experiments, and our claims about AAP remain hypotheses until validated. Our toy illustrations (e.g., the Q→U spelling rule) capture asymmetry in simplified form but do not represent the full complexity of natural language. Moreover, the proposed objectives may introduce optimization trade-offs such as higher computational cost from evaluating reversed sequences, or stability challenges from hinge-style penalties. Future work will need to explore these trade-offs and test AAP on morphologically rich languages, multimodal corpora, and long-range discourse. We emphasize that our intent is to lay a conceptual foundation and to motivate empirical inquiry; validation and practical results are left for subsequent work.

11 Ethics Statement

This work is purely theoretical and does not involve sensitive data collection or human subjects. Its intended contribution is to clarify conceptual limits of current reasoning claims in LLMs and to motivate new lines of empirical inquiry. If asymmetry-aware training were validated in future work, it could contribute to developing models that are more transparent and reliable in reasoning contexts.

Acknowledgements

The author thanks colleagues and mentors for discussion and feedback. Any errors or oversights remain the sole responsibility of the author.

References

[1] Lukas Berglund, Meg Tong, Max Kaufmann, Mikita Balesni, Asa Cooper Stickland, Tomasz Korbak, and Owain Evans. The reversal curse: Llms trained on "a is b" fail to learn "b is a", 2024.

[2] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

[3] Paul Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences, 2023.

[4] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, 2006.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[6] Olga Golovneva, Zeyuan Allen-Zhu, Jason Weston, and Sainbayar Sukhbaatar. Reverse training to nurse the reversal curse. *arXiv preprint arXiv:2403.13799*, 2024.

[7] Qingyan Guo, Rui Wang, Junliang Guo, Xu Tan, Jiang Bian, and Yujiu Yang. Mitigating reversal curse in large language models via semantic-aware permutation training. *arXiv preprint arXiv:2403.00758*, 2024.

[8] Licheng Jiao, Yuhan Wang, Xu Liu, Lingling Li, Fang Liu, Wenping Ma, Yuwei Guo, Puhua Chen, Shuyuan Yang, and Biao Hou. Causal inference meets deep learning: A comprehensive survey. *Research*, 7:0467, 2024.

[9] Sékou-Oumar Kaba and Siamak Ravanbakhsh. Symmetry breaking and equivariant neural networks, 2024.

[10] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners, 2023.

[11] Christian Maes and Karel Netočný. Time-reversal and entropy. *Journal of Statistical Physics*, 110(1-2):269–310, 2003.

[12] Samuel E. Otto, Nicholas Zolman, J. Nathan Kutz, and Steven L. Brunton. A unified framework to enforce, discover, and promote symmetry in machine learning, 2025.

[13] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback, 2022.

[14] Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. Causal inference using invariant prediction: identification and confidence intervals. In *Journal of the Royal Statistical Society Series B*, volume 78, page 947–1012, 2016.

[15] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In *Proceedings of NAACL-HLT*, pages 2227–2237, 2018.

[16] Maxime Peyrard. Invariant language modeling. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2022.

[17] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. *arXiv preprint arXiv:1801.06146*, 2018.

[18] Raanan Y. Rohekar, Yaniv Gurwicz, and Shami Nisimov. Causal interpretation of self-attention in pre-trained transformers. In *NeurIPS 2023 Poster Session*, 2023. Poster, NeurIPS 2023.

[19] Matteo Saponati, Pascal Sager, Pau Vilimelis Aceituno, Thilo Stadelmann, and Benjamin Grewe. The underlying structures of self-attention: symmetry, directionality, and emergent dynamics in transformer training, 2025.

[20] Yair Schiff, Chia-Hsiang Kao, Aaron Gokaslan, Tri Dao, Albert Gu, and Volodymyr Kuleshov. Caduceus: Bi-directional equivariant long-range dna sequence modeling. *arXiv preprint arXiv:2403.03234*, 2024.

[21] Rico Sennrich, Barry Haddow, and Alexandra Birch. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016.

[22] Zhaochen Su, Jun Zhang, Tong Zhu, Xiaoye Qu, Juntao Li, Min Zhang, and Yu Cheng. Timo: Towards better temporal reasoning for language models, 2024.

[23] Hidenori Tanaka and Daniel Kunin. Noether's learning dynamics: Role of symmetry breaking in neural networks, 2021.

[24] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

[25] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023.

[26] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Pan Pasupat, and Quoc Le. Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*, 2022.

[27] Xinyi Wu, Amir Ajorlou, Yifei Wang, Stefanie Jegelka, and Ali Jadbabaie. On the role of attention masks and layernorm in transformers. In *NeurIPS 2024 Poster Session*, 2024. Poster, NeurIPS 2024.

[28] Siheng Xiong, Ali Payani, Ramana Kompella, and Faramarz Fekri. Large language models can learn temporal reasoning, 2024.

[29] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

[30] Sicheng Yu, Yuanchen Xu, Cunxiao Du, Yanying Zhou, Minghui Qiu, Qianru Sun, Hao Zhang, and Jiawei Wu. Reverse modeling in large language models, 2025.

[31] Jianyuan Zhang, Deyi Xiong, Feng Wei, and Zhongwen He. Regularizing right-to-left models for neural machine translation. *Transactions of the Association for Computational Linguistics*, 7:353–365, 2019.

[32] Jiaqi Zhang, Joel Jennings, Agrin Hilmkil, Nick Pawlowski, Cheng Zhang, and Chao Ma. Towards causal foundation model: on duality between causal inference and attention, 2024.

[33] Shiyue Zhang, Shijie Wu, Ozan Irsoy, Steven Lu, Mohit Bansal, Mark Dredze, and David Rosenberg. Mixce: Training autoregressive language models by mixing forward and reverse cross-entropies. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.

[34] Bo Zhao, Robin Walters, and Rose Yu. Symmetry in neural network parameter spaces, 2025.