DEPTH-LINKED REGISTER TRANSFORMER

Anonymous authors
Paper under double-blind review

ABSTRACT

Looped language models (LoopLMs) have emerged as a powerful paradigm for improving parameter efficiency while retaining strong reasoning capacity compared to fixed-depth LLMs. However, their reliance on sequential, autoregressive iterations incurs substantial computational overhead. In this work, we revisit parallelizable register embeddings for non-autoregressive latent reasoning. A key challenge of this paradigm, however, is the absence of iterative refinement and the resulting weaker sequential dependencies between latent thoughts. To address these limitations, we propose the **Depth**-L**ink**ed Register Transformer (**DepThink**), a non-autoregressive latent reasoning framework that simulates the sequential dependency and representational expressiveness of LoopLMs while preserving the efficiency of parallel inference. DepThink employs a sequence of K register embeddings within a single pre-filling stage, where each register captures information from the input that is missed by its preceding registers through learnable residual connections, yielding a progressively refined latent thought. This design preserves the iterative refinement nature of the thinking steps intrinsic to looped reasoning, while inheriting the parallelism and representational capacity of register-based models, and eliminating the K sequential backbone passes required by LoopLMs. Moreover, we theoretically prove that, relative to the standard transformer with registers, DepThink achieves stronger geometric directional reachability and a provable one-step descent advantage. We evaluate DepThink on reasoning-intensive text generation tasks, such as mathematical reasoning and knowledge-rich reasoning, as well as reasoning-intensive multimodal retrieval. Experimental results demonstrate that DepThink outperforms standard looped architectures and Transformers with registers across dense (2B, 8B) and sparse mixture-of-experts (30B-A3B) backbones, under both LoRA and full-parameter fine-tuning. Furthermore, wall-clock measurements reveal that DepThink can accelerate training by $11.9\times$ and inference by $4.3\times$ compared to standard LoopLMs.

1 INTRODUCTION

The evolution of Large Language Models (LLMs) has recently shifted from a pure focus on increasing parameter counts to maximizing the computational utility of existing weights. Among the most promising developments in this direction are looped architectures (Yang et al., 2023; Xu & Sato, 2024; Saunshi et al., 2025). By sharing or reusing parameters across the depth of the network, these models achieve remarkable parameter efficiency and demonstrate an improved capacity for complex reasoning. They effectively allow a model to "think longer" by iterating through the same layers multiple times, aiming to achieve similar capacity of LLMs with equivalent depth of layers.

However, iterative approaches, especially those loop entire transformer layers (Hao et al., 2024; Zhu et al., 2025), incur substantial computational cost, since each step must wait for the previous one to complete due to autoregressive dependencies. Non-autoregressive latent reasoning (Goyal et al., 2024; Su et al., 2025) avoids this sequential bottleneck by reusing shared layers across iterations, but at the cost of representational capacity and weak sequential dependencies among intermediate thinking states. In this work, we revisit the use of parallelizable register tokens, originally introduced in vision transformers (Darcet et al., 2023), as a tool for non-autoregressive latent reasoning that preserves computational efficiency while addressing the absence of iterative refinement inherent to such designs by **residual connections** between thinking states.

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.

More specifically, we introduce **Depth**-L**ink**ed Register Transformer (**DepThink**), a non-autoregressive latent reasoning model that achieves both computational efficiency and strong reasoning and representational capacity. At its core, DepThink leverages learnable register embeddings appended to the input sequence, linking their hidden states through explicit depth-wise residual connections. Rather than recursively looping the entire network, each register's hidden state extracts and injects an incremental piece of information into a progressively cumulative latent thought. These lightweight residual updates couple different register positions into an ordered latent-thought trajectory, and simulate the iterative refinement of looped language models (LoopLMs) without incurring their autoregressive cost or sacrificing per-step representational diversity. A key advantage of this design is the shift of computational workload from sequential backbone passes to the highly parallelizable pre-filling stage. By processing all register embeddings within a single forward pass, DepThink computes the "reasoning residuals" simultaneously, thereby reducing reasoning latency significantly when compared to LoopLMs.

Beyond these computational efficiencies, our rigorous theoretical analysis establishes that DepThink provably overcomes the geometric constraints of standard registers, guaranteeing enhanced directional reachability and a strict first-order optimization advantage. We evaluate DepThink on both text generation and multi-modal retrieval tasks, including math reasoning datasets GSM8K (Cobbe et al., 2021) and MATH (Hendrycks et al., 2021), knowledge question answering datasets Natural Questions (NQ) (Kwiatkowski et al., 2019) and HotpotQA (Yang et al., 2018), and the multi-modal retrieval datasets CIRR (Liu et al., 2021), MSCOCO (Lin et al., 2014), MSVD (Chen & Dolan, 2011), and QVHighlights (Lei et al., 2021). We further evaluate full-parameter fine-tuning, dense 8B scaling, and a sparse 30B-A3B mixture-of-experts backbone. Across these settings, DepThink is competitive with or better than LoopLM and Register while requiring only a single backbone pass for its latent thoughts. By treating latent thoughts as an incremental accumulation of information guided by DepThink registers, we provide a path toward models that possess the depth and logic of LoopLMs without the associated temporal and computational costs.

2 Related Work

Looped Transformers Reusing parameters across the depth of a model is a long-standing strategy for improving model efficiency. The Universal Transformer (Dehghani et al., 2019) introduced the idea for sequence models, which has recently regained interest to shrink memory footprints for edge deployment (Li et al., 2025), enable test-time scaling (Wu et al., 2025; Geiping et al., 2025) and language modeling with latent reasoning (Zhu et al., 2025). Csordás et al. (2021) generalize Universal Transformers to MoE architecture; Saunshi et al. (2025) show that looped models excel at reasoning; Fan et al. (2024) demonstrate length generalization with looped transformers. Prairie et al. (2026) study scaling laws for loop-structure. Zeitoun et al. (2026) combines hyper-connections with looped models to further improve the parameter efficiency. Despite its success, the efficiency of looped transformer is limited by its sequential nature, which motivates our work.

Latent Reasoning Chain-of-Thought (CoT) (Wei et al., 2022; Nye et al., 2021) has become the dominant paradigm for eliciting reasoning in LLMs, but verbalizing every reasoning step inflates inference cost and may not be the most efficient computational medium for LLMs (Deng et al., 2023; Guo et al., 2025). Recent work therefore moves reasoning into the model's latent space (Deng et al., 2023; 2024; Hao et al., 2024). Broadly, looped and recurrent architectures enable arbitrary-depth reasoning without emitting tokens (Dehghani et al., 2018; Yang et al., 2023; Geiping et al., 2025; Saunshi et al., 2025). Beyond autoregressive models, non-autoregressive formulations has been explored with diffusion-based latent reasoning (Ye et al., 2024). Orthogonally, special-purpose tokens such as pause/filler tokens (Goyal et al., 2024; Pfau et al., 2024) and planning tokens (Wang et al., 2023b) have also been used to support latent reasoning.

Learnable Special Tokens Another line of research augments transformer inputs with small sets of learnable tokens not in the pre-trained vocabulary. In NLP, soft prompt tuning (Lester et al., 2021) and prefix tuning (Li & Liang, 2021) prepend a sequence of trainable embeddings to the input of a frozen language model for parameter efficient training. In computer vision, register tokens are widely used in Vision Transformers (Dosovitskiy et al., 2020) for global feature learning (Darcet et al., 2023). In multimodal learning, CoOp (Zhou et al., 2022b) and CoCoOp (Zhou et al., 2022a) proposes to

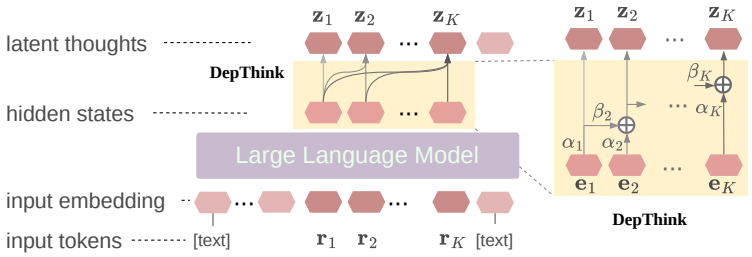


Figure 1: **Overview of DepThink model.** Given the input tokens, DepThink appends K learnable register embeddings $\{\mathbf{r}_1, \ldots, \mathbf{r}_K\}$. Their hidden states $\{\mathbf{e}_1, \ldots, \mathbf{e}_K\}$ are linked by a learnable residual connection to produce K progressively refined latent thoughts $\{\mathbf{z}_1, \ldots, \mathbf{z}_K\}$.

optimize the language context for vision-language alignment. BLIP-2's Q-Former (Li et al., 2023a) uses a fixed set of learnable queries to extract a compact, language-aligned representation from a frozen visual encoder via cross-attention; the resulting query outputs serve as a soft visual prompt to a frozen LLM. The pattern was also adopted by InstructBLIP (Dai et al., 2023) and Flamingo's Perceiver Resampler (Alayrac et al., 2022) as a modality bridge.

Learnable Residual Connections Learnable residual mechanisms are traditionally used to optimize very deep networks. For example, Highway Networks (Srivastava et al., 2015) gate information across layers, ReZero (Bachlechner et al., 2021) accelerates convergence with zero-initialized gates, and LayerScale (Touvron et al., 2021) stabilizes deep vision transformers through per-channel scaling. Although DepThink utilizes a similar algebraic primitive, its objective is fundamentally different. Instead of scaling the layer-wise token residual stream, it couples distinct register positions to generate a sequence of latent thoughts in a single forward pass for LLMs. Ultimately, DepThink's novelty lies not in merely introducing a learnable residual coefficient, but in how this recurrence approximates loop-style refinement and enhances register reachability.

3 Approach

3.1 Register Transformer and DepThink Models

Let $T_{\boldsymbol{\theta}} : \mathbb{R}^{(J+K)\times D} \to \mathbb{R}^{(J+K)\times D}$ be a causal transformer backbone (e.g., a LoRA-adapted Qwen3-VL backbone) with parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The backbone consumes a sequence of $J + K$ embeddings and produces a contextualized sequence of the same length. For any input $x \in \mathcal{X}$ where $\mathcal{X}$ is the input space of queries and targets, let $\mathrm{Emb}(x) \in \mathbb{R}^{J\times D}$ denote the corresponding input-token embeddings. We introduce a set of learnable register embeddings $\mathbf{R} = [\mathbf{r}_1; \ldots; \mathbf{r}_K] \in \mathbb{R}^{K\times D}$ which are appended to the input sequence $\mathrm{Emb}(x)$. The forward pass can be expressed as

$$\big[\mathbf{h}_1(x); \ldots; \mathbf{h}_J(x);\ \mathbf{e}_1(x); \ldots; \mathbf{e}_K(x)\big] \ \triangleq\ T_{\boldsymbol{\theta}}\big([\mathrm{Emb}(x);\ \mathbf{r}_1;\ \ldots;\ \mathbf{r}_K]\big). \tag{1}$$

In (1), the first J outputs $\{\mathbf{h}_j(x)\}_{j=1}^{J}$ represent the hidden states of the input tokens, while the subsequent K outputs $\{\mathbf{e}_k(x)\}_{k=1}^{K} \subset \mathbb{R}^D$ denote the hidden states of the registers. We refer to $\mathbf{e}_k(x)$ as the hidden state of the k-th register. Under the standard causal attention mask, $\mathbf{e}_k$ depends on $\mathrm{Emb}(x)$ and on the learnable registers $\mathbf{r}_1, \ldots, \mathbf{r}_k$. We collect them into a multi-vector $\mathbf{E}(x) \triangleq [\mathbf{e}_1(x); \ldots; \mathbf{e}_K(x)] \in \mathbb{R}^{K\times D}$. For the standard Register Transformer (referred to as Register hereafter), the output is the identity mapping on $\mathbf{E}$, that is,

$$\boldsymbol{\Phi}^{\lambda}(x) \ \triangleq\ \mathbf{E}(x) = [\mathbf{e}_1(x); \ldots; \mathbf{e}_K(x)] \ \in\ \mathbb{R}^{K\times D}. \tag{2}$$

In contrast, as illustrated in Fig. 1, DepThink transforms the multi-vector $\mathbf{E}(x)$ by integrating a learnable first-order linear Infinite Impulse Response (**IIR**) recurrence, effectively linking registers' hidden states across the depth dimension. Specifically, let $\boldsymbol{\alpha} \triangleq [\alpha_1, \ldots, \alpha_K] \in \mathbb{R}^K$ and $\boldsymbol{\beta} \triangleq [0, \beta_2, \ldots, \beta_K] \in \mathbb{R}^K$ denote the learnable scaling and recurrence coefficients, respectively. The output of the k-th register in DepThink, which is called a **latent thought**, is then given by

$$\mathbf{z}_k(x) \ \triangleq\ \beta_k\, \mathbf{z}_{k-1}(x) \ +\ \alpha_k\, \mathbf{e}_k(x), \qquad \mathbf{z}_0(x) \triangleq 0. \tag{3}$$

Unrolling the recurrence in (3) yields the following closed-form expression

$$\mathbf{z}_k(x) \;=\; \sum_{i=1}^{k} M_{ki}(\boldsymbol{\alpha},\boldsymbol{\beta})\,\mathbf{e}_i(x), \qquad M_{ki}(\boldsymbol{\alpha},\boldsymbol{\beta}) \;\triangleq\; \alpha_i \prod_{j=i+1}^{k} \beta_j, \tag{4}$$

where we adopt the convention that the empty product equals 1 when $i = k$, and hence $M_{kk} = \alpha_k$. The transformation is characterized by the lower-triangular matrix $\mathbf{M}(\boldsymbol{\alpha},\boldsymbol{\beta}) \in \mathbb{R}^{K\times K}$ which possesses full rank provided $\alpha_k \neq 0$ for all k. Consequently, the DepThink head can be expressed as the concatenation of these refined latent thoughts

$$\boldsymbol{\Phi}^{\mathrm{dep}}(x) \triangleq [\mathbf{z}_1(x);\ldots;\mathbf{z}_K(x)] \in \mathbb{R}^{K\times D}. \tag{5}$$

For each depth k, we define the unit-norm vectors $\widehat{\mathbf{z}}_k(x)$ and $\widehat{\mathbf{e}}_k(x)$ to project the representations onto the unit sphere $\mathbb{S}^{D-1} \triangleq \{\mathbf{u} \in \mathbb{R}^D : \|\mathbf{u}\|_2 = 1\}$, that is,

$$\widehat{\mathbf{z}}_k(x) \;\triangleq\; \frac{\mathbf{z}_k(x)}{\|\mathbf{z}_k(x)\|_2} \;\in\; \mathbb{S}^{D-1} \quad \text{and} \quad \widehat{\mathbf{e}}_k(x) \;\triangleq\; \frac{\mathbf{e}_k(x)}{\|\mathbf{e}_k(x)\|_2} \;\in\; \mathbb{S}^{D-1}. \tag{6}$$

We assume $\|\mathbf{z}_k(x)\|_2 > 0$ over the support of the input distribution. This naturally holds for non-degenerate encoders, as the set of inputs yielding a zero latent thought has Lebesgue measure zero. Finally, we define the pairwise similarity score as the average per-depth cosine similarity across matched depth indices. For DepThink and Register, their pairwise similarity scores can be written as:

$$S^{\mathrm{dep}}(x,y) \;\triangleq\; \frac{1}{K}\sum_{k=1}^{K}\langle \widehat{\mathbf{z}}_k(x), \widehat{\mathbf{z}}_k(y)\rangle \quad \text{and} \quad S^{\lambda}(x,y) \;\triangleq\; \frac{1}{K}\sum_{k=1}^{K}\langle \widehat{\mathbf{e}}_k(x), \widehat{\mathbf{e}}_k(y)\rangle. \tag{7}$$

3.2 Training Objectives

DepThink learns a sequence of latent thoughts $\{\mathbf{z}_k\}_{k=1}^{K}$ and supports both text generation and representation learning. For text generation, the depth-linked registers forms a latent reasoning trajectory that matches the functional use of "latent thought" in prior literature. For representation learning, the latent thoughts encode the input together with its instruction and are evaluated directly through multimodal retrieval.

Conditional Text Generation Given a prompt x_i, we generate target text conditioned on $\boldsymbol{\Phi}^{\mathrm{dep}}(x)$ as defined in (5). Formally, consider a batch of N input samples $\{(x_i,u_i)\}_{i=1}^{N}$, where $u_i = (u_i^0, u_i^1, \ldots, u_i^m)$ denotes the target sequence corresponding to input tokens x_i. Each token u_i^j is generated auto-regressively, conditioned on the prefix tokens $u_i^{<j}$, and the register latent thoughts $\mathbf{z}_k(x_i)$. We optimize the parameters via the standard negative log-likelihood auto-regressive loss $\mathcal{L}_{\mathrm{AR}}(\boldsymbol{\Phi},\boldsymbol{\theta}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{m}\log P\left(u_i^j \mid u_i^{<j},\, \boldsymbol{\Phi}^{\mathrm{dep}}(x_i);\, \boldsymbol{\theta}\right)$.

Representation Learning To train DepThink for retrieval and alignment, we employ a contrastive objective. Given a batch of N positive pairs $\{(x_i,y_i)\}_{i=1}^{N}$ drawn independently and identically distributed (i.i.d.) from a joint distribution $\mathcal{P}$ on $\mathcal{X}\times\mathcal{X}$, and a temperature parameter $\tau > 0$, the empirical InfoNCE (Oord et al., 2018) risk for a head $\boldsymbol{\Phi} \in \{\boldsymbol{\Phi}^{\lambda}, \boldsymbol{\Phi}^{\mathrm{dep}}\}$ with corresponding similarity score $S \in \{S^{\lambda}, S^{\mathrm{dep}}\}$ is defined as:

$$\hat{L}_N(\boldsymbol{\Phi}) \;\triangleq\; -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(S(x_i,y_i)/\tau)}{\sum_{j=1}^{N}\exp(S(x_i,y_j)/\tau)}. \tag{8}$$

The corresponding population InfoNCE risk is then given by the expectation over the data distribution

$$L_{\mathcal{P}}(\boldsymbol{\Phi}) \;\triangleq\; \mathbb{E}_{(x_i,y_i)\sim\mathcal{P}^N}\big[\hat{L}_N(\boldsymbol{\Phi})\big]. \tag{9}$$

Superiority to Register: In conditional text generation, reasoning-intensive generation needs step-by-step deduction, but the unlinked registers in Register interact only through causal attention, so they tend to stay isolated rather than form a coherent reasoning chain. DepThink chains the registers through a learnable recurrence during pre-filling, which gives it the iterative refinement of looped

models without their autoregressive latency (Giannou et al., 2023) and builds an implicit reasoning chain that improves generation quality. In representation learning tasks such as cross-modal alignment, Register's unlinked registers tend to encode redundant information. DepThink instead accumulates the representation step by step: each register is trained to capture what the earlier registers missed. As shown below, this lets the registers reach directions that unlinked registers cannot, which richer semantic aggregation needs.

3.3 Computational Efficiency

One of the main advantages of DepThink and Register models over LoopLMs is the computational efficiency. Specifically, all K latent thoughts are computed in one forward pass over $L+K$ tokens, whereas a LoopLM performs K sequential forward passes. Let $C_b = \mathcal{O}(LD^2 + L^2D)$ denote the baseline cost of a vanilla Transformer pass over L tokens of width D, accounting for projections/feed-forward layers and attention. For the standard LoopLM, it performs K sequential backbone passes that leads to $C_l = K \cdot C_b$. However, the DepThink and Registers models perform one pass over $L+K$ tokens such that the cost is $C_r = \mathcal{O}((L+K)D^2 + (L+K)^2D)$, the ratio of which over Transformer baseline is within $(1+K/L)$ and $(1+K/L)^2$. In practice, $K << L$ due to the long content context of content tokens such that $(1+K/L)^2$ is close to 1.0, indicating very marginal computation overhead. In Table 6, we reported efficiency benchmarks that validated this analysis.

4 Theoretical analysis

4.1 Lower InfoNCE Risk Guarantee via Subsumption: Register $\subseteq$ DepThink

For DepThink and Register, we define two classes for their heads, respectively: $\mathcal{H}_\lambda \triangleq \{\mathbf{\Phi}^\lambda : (\boldsymbol{\theta}, \mathbf{R}) \in \mathbf{\Theta} \times \mathbb{R}^{K\times D}\}$ and $\mathcal{H}_{\text{dep}} \triangleq \{\mathbf{\Phi}^{\text{dep}} : (\boldsymbol{\theta}, \mathbf{R}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathbf{\Theta} \times \mathbb{R}^{K\times D} \times \mathbb{R}^K \times \mathbb{R}^K\}$. Let $\boldsymbol{\eta} \triangleq (\boldsymbol{\theta}, \mathbf{R}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ denote the full parameter tuple for the DepThink head. To bridge these two spaces, we introduce a lift map, which embeds the Register parameter space into that of DepThink

$$\boldsymbol{\eta}_0(\boldsymbol{\theta}, \mathbf{R}) \triangleq (\boldsymbol{\theta}, \mathbf{R}, \mathbf{1}, \mathbf{0}). \tag{10}$$

Furthermore, we define the optimal population risk for a given class $\mathcal{H}$ at sample size N under distribution $\mathcal{P}$ as

$$R^*(\mathcal{H}, \mathcal{P}) \triangleq \inf_{\mathbf{\Phi}\in\mathcal{H}} L_{\mathcal{P}}(\mathbf{\Phi}). \tag{11}$$

Proposition 1 (Lower InfoNCE Risk via Subsumption)**.** *For any backbone configuration $(\boldsymbol{\theta}, \mathbf{R}) \in \mathbf{\Theta} \times \mathbb{R}^{K\times D}$, we have $\mathbf{\Phi}^{\text{dep}}_{\boldsymbol{\eta}_0(\boldsymbol{\theta},\mathbf{R})} = \mathbf{\Phi}^{\lambda}_{(\boldsymbol{\theta},\mathbf{R})}$ as functions from $\mathcal{X} \to \mathbb{R}^{K\times D}$. Consequently, $\mathcal{H}_\lambda \subseteq \mathcal{H}_{\text{dep}}$, which implies $R^*(\mathcal{H}_{\text{dep}}, \mathcal{P}) \le R^*(\mathcal{H}_\lambda, \mathcal{P})$ for any distribution $\mathcal{P}$ and sample size N.*

The proof of Proposition 1 is deferred to Appendix A. Proposition 1 ensures that DepThink strictly subsumes the standard Register architecture. By providing a distortion-free embedding ($\boldsymbol{\eta}_0$) of any Register checkpoint, it guarantees that DepThink can match any Register-attainable performance without incurring an architectural penalty. Crucially, this subsumption is dimensionally non-trivial: while Register is fixed to a singular direction, DepThink unlocks a higher-dimensional family of directional choices per input, as formalized in the subsequent analysis.

4.2 Geometric Advantage: Enhanced Directional Reachability of DepThink

Fix the backbone parameters $(\boldsymbol{\theta}, \mathbf{R})$ and an input $x \in \mathcal{X}$. We define the depth-k Register span and its corresponding rank as

$$\mathbb{V}_k(x) \triangleq \text{span}\{\mathbf{e}_1(x), \ldots, \mathbf{e}_k(x)\} \subseteq \mathbb{R}^D \quad \text{and} \quad d_k(x) \triangleq \dim \mathbb{V}_k(x) \in \{1, \ldots, k\}.$$

Holding the tuple $(\boldsymbol{\theta}, \mathbf{R}, x)$ fixed and varying the IIR coefficients $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, the per-input depth-$k$ reachable direction set for DepThink is formally defined as

$$\mathcal{D}^{\text{dep}}_k(x; \boldsymbol{\theta}, \mathbf{R}) \triangleq \big\{\widehat{\mathbf{z}}_k(x; \boldsymbol{\alpha}, \boldsymbol{\beta}) : (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathbb{R}^K \times \mathbb{R}^K,\ \|\mathbf{z}_k(x; \boldsymbol{\alpha}, \boldsymbol{\beta})\|_2 > 0\big\}.$$

Theorem 2 (Directional Reachability)**.** *For any backbone configuration $(\boldsymbol{\theta}, \mathbf{R})$, any input x such that $\mathbb{V}_k(x) \neq \{\mathbf{0}\}$, and any depth $k \in \{1, \ldots, K\}$, we have*

$$\mathcal{D}^{\text{dep}}_k(x; \boldsymbol{\theta}, \mathbf{R}) = \mathbb{V}_k(x) \cap \mathbb{S}^{D-1}. \tag{12}$$

In contrast, the reachable set for Register is strictly the singleton $\mathcal{D}^{\lambda}_k(x) = \{\widehat{\mathbf{e}}_k(x)\}$.

The proof of Theorem 2 is provided in Appendix B. Theorem 2 formalizes the core representational advantage intrinsic to DepThink: a strict expansion of per-input directional freedom. Concretely, for any backbone configuration $(\boldsymbol{\theta}, \mathbf{R})$ and any non-degenerate input x, the IIR coefficients inject exactly $(d_k(x) - 1)$ additional degrees of freedom into the depth-k representation. As the coefficients $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ vary, the DepThink direction $\widehat{\mathbf{z}}_k$ spans the entire $(d_k(x)-1)$-dimensional hypersphere $\mathbb{V}_k(x) \cap \mathbb{S}^{D-1}$. In contrast, the standard Register remains rigidly anchored to a zero-dimensional point $\{\widehat{\mathbf{e}}_k(x)\}$.

4.3 Optimization Advantage: Provable One-Step Descent Superiority

We now demonstrate that the enhanced directional reachability of DepThink established in Theorem 2 translates directly into a provable first-order optimization advantage over the Register. Specifically, we analyze a single gradient descent step originating from an arbitrary Register checkpoint, which is embedded into the DepThink parameter space via $\boldsymbol{\eta}_0$ in (10).

Under a local smoothness assumption on the expected InfoNCE risk $L_{\mathcal{P}}$ (formalized in Assumption 1 in Appendix C), an analysis of the partial derivatives at the lift point $\boldsymbol{\eta}_0$ reveals a gradient decoupling. Specifically, because of the scale-invariance inherent to the ℓ_2-normalization, the gradient with respect to the scaling parameters $\boldsymbol{\alpha}$ vanishes identically. In contrast, the gradient with respect to the recurrence coefficients $\boldsymbol{\beta}$ remains active, effectively capturing the projection of the preceding depth's hidden state onto the orthogonal complement of the current hidden state (see Lemma 5 and Theorem 6 in Appendix C for closed-form derivations). The squared norm of this non-vanishing component, $\|\nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0}\|^2$, governs the enhanced risk reduction during the gradient step.

Let $(\boldsymbol{\theta}', \mathbf{R}') \triangleq (\boldsymbol{\theta}, \mathbf{R}) - \mu \mathbf{g}^{\lambda}$ and $\boldsymbol{\eta}' \triangleq \boldsymbol{\eta}_0 - \mu \mathbf{g}^{\mathrm{dep}}$ denote the updated parameter configurations following a single gradient descent step with step size $\mu > 0$, where $\mathbf{g}^{\lambda}$ and $\mathbf{g}^{\mathrm{dep}}$ represent the gradients of the expected InfoNCE risk for Register and DepThink, respectively. We define the expected InfoNCE risk gap between Register and DepThink after the single descent step as

$$\Delta(\mu) \;\triangleq\; L_{\mathcal{P}}\big(\boldsymbol{\Phi}^{\lambda}_{(\boldsymbol{\theta}', \mathbf{R}')}\big) - L_{\mathcal{P}}\big(\boldsymbol{\Phi}^{\mathrm{dep}}_{\boldsymbol{\eta}'}\big). \tag{13}$$

Lemma 3 (Descent Gap Expansion)**.** *Under the local smoothness Assumption 1 in Appendix C, for any sufficiently small $\mu > 0$, the expected InfoNCE risk gap admits the exact closed-form expression*

$$\Delta(\mu) \;=\; \mu \big\| \nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} \big\|^2 + O(\mu^2). \tag{14}$$

The proof of Lemma 3 is provided in Appendix C, which also establishes an upper bound on the remainder term in (14). This expansion implies that, provided the $\boldsymbol{\beta}$-gradient is non-zero, the first-order linear descent term strictly dominates the local quadratic curvature penalty for sufficiently small μ. By bounding this curvature via the local Lipschitz constant L_{loc} of the risk gradient (defined in Assumption 1 in Appendix C), we can analytically determine the exact step-size boundary that guarantees a greater reduction in the expected risk. We summarize these results in the following theorem, whose detailed proof is deferred to Appendix C.

Theorem 4 (Weak and Strict Optimization Advantage)**.** *Suppose the expected InfoNCE risk satisfies the local smoothness Assumption 1 in Appendix C with Lipschitz constant $L_{\mathrm{loc}} > 0$, admitting a maximum valid step size $\mu_{\max} > 0$ that confines the update to the locally Lipschitz neighborhood. Define the critical step-size threshold*

$$\bar{\mu} \;\triangleq\; \frac{2\big\| \nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} \big\|^2}{L_{\mathrm{loc}} \left(\|\mathbf{g}^{\lambda}\|^2 + \|\mathbf{g}^{\mathrm{dep}}\|^2 \right)}. \tag{15}$$

Then, for any gradient descent step size $\mu \in \big(0, \min(\bar{\mu}, \mu_{\max})\big)$, the following properties hold

1. ***Weak Optimization Advantage:*** *$\Delta(\mu) \geq 0$ whenever $\mu \in \big(0, \min(\bar{\mu}, \mu_{\max})\big)$. Furthermore, the exact equality $\Delta(\mu) = 0$ holds if and only if $\nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} = \mathbf{0}$, in which case the descent trajectories of DepThink and Register perfectly coincide.*
2. ***Strict Optimization Advantage:*** *If $\nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} \neq \mathbf{0}$, then $\Delta(\mu) > 0$ strictly for all $\mu \in \big(0, \min(\bar{\mu}, \mu_{\max})\big)$, guaranteeing that DepThink consistently achieves a strictly greater InfoNCE risk reduction than Register.*

The strict optimization advantage established in Theorem 4 hinges on a non-vanishing first-order signal, specifically $\nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} \neq \mathbf{0}$. From an analytical perspective, this condition holds

Table 1: GSM8K accuracy. † denotes our controlled 2B re-implementation.

Method	Size	Accuracy (%)
Pause Token (Goyal et al., 2024)	3B	53.70
Token Assorted (Su et al., 2025)	1B	48.70
Coconut (Hao et al., 2024)	≤1.5B	34.10
Coconut†	2B	38.04
Register	2B	68.29
LoopLM	2B	70.12
DepThink	2B	**71.11**

Table 2: MATH and MATH-500 accuracy (%).† denotes our controlled 2B re-implementation.

Method	Size	MATH	MATH-500
HRPO (Yue et al., 2026)	3B	61.30	63.30
CRISP (Li et al., 2026)	3B	–	76.00
Coconut†	2B	71.30	71.00
LoopLM	2B	81.72	79.80
Register	2B	82.50	81.60
DepThink	2B	**82.54**	**82.40**

Lebesgue almost everywhere across the parameter space. This is because modern transformer architectures equipped with smooth activations usually formulate a real-analytic gradient map under the InfoNCE objective. Given that real-world multimodal data distributions inherently violate rigid group symmetries (such as absolute continuous isotropy and exact query-target exchangeability), the expected InfoNCE gradient map cannot vanish identically. By the fundamental properties of real-analytic functions, its zero set must therefore be strictly confined to a subset of Lebesgue measure zero (Mityagin, 2015). Consequently, any random parameter initialization or checkpoint encountered during stochastic optimization will almost surely avoid this zero set, robustly guaranteeing the strict optimization advantage in Theorem 4 in general.

5 Empirical Study

Experimental setup. We evaluate DepThink in reasoning-intensive applications including the math reasoning, knowledge question answering, and multimodal retrieval tasks. For math reasoning, we use GSM8K (Cobbe et al., 2021), MATH (Hendrycks et al., 2021), and MATH-500 (Hendrycks et al., 2021) datasets. For knowledge QA, we evaluate models on Natural Questions (NQ) (Kwiatkowski et al., 2019) and HotpotQA (Yang et al., 2018) datasets. For multimodal retrieval, we compare against existing approaches on the CIRR (Liu et al., 2021), MSCOCO (Lin et al., 2014), MSVD (Chen & Dolan, 2011), and QVHighlights (Lei et al., 2021), which evaluate cross-modality retrieval among text, image, and video.

Table 3: NQ and HotpotQA results evaluated by Exact Match (EM) scores.

Method	NQ	HotpotQA		
		Direct	Think	Retrieve
Register	16.12	17.32	15.61	32.55
LoopLM	16.15	**17.65**	15.49	16.81
DepThink	**16.57**	17.43	**15.84**	**32.65**

Implementation details. We implement the DepThink primarily on the Qwen3-VL-Instruct 2B by default, that supports both text-only and multi-modality tasks. We use LoRA adapter of rank 64 for different tasks . We initialize $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to 1 and 0, respectively, so that DepThink starts from the Register solution. We use $K = 4$ registers for retrieval and $K = 16$ for text generation by default, and performed comprehensive analysis. Unless stated otherwise, training uses AdamW with learning rate 2×10^{-4}, weight decay 0.1, a cosine schedule with 500 warm-up steps, 20,000 updates on 32 H100 GPUs, and an effective batch size of 1,024.

5.1 Main Results

Math reasoning. In this task, we first compare DepThink against recent latent reasoning approaches, Coconut (Hao et al., 2024), Pause Token (Goyal et al., 2024) and Token Assorted (Su et al., 2025). In this track, we train models on GSM8K (Cobbe et al., 2021) training splits and evaluate on its test split. We set $K = 16$ as the number of registers for DepThink and Registers and comparatively set the loop depth and reasoning depth to 16 for LoopLM and Coconut models. Table 1 shows that DepThink reaches to the best performance at 71.11%. The relatively poor performance of Coconut can be attributed to its direct answer with implicit reasoning, while the Register, LoopLM, and DepThink explicitly write out reasoning traces. In addition, following (Yue et al., 2026), we also train models with $K = 16$ on MATH training split and evaluate on both MATH and MATH-500 test data. Table 2 further validate the superiority of our DepThink method on math problem solving.

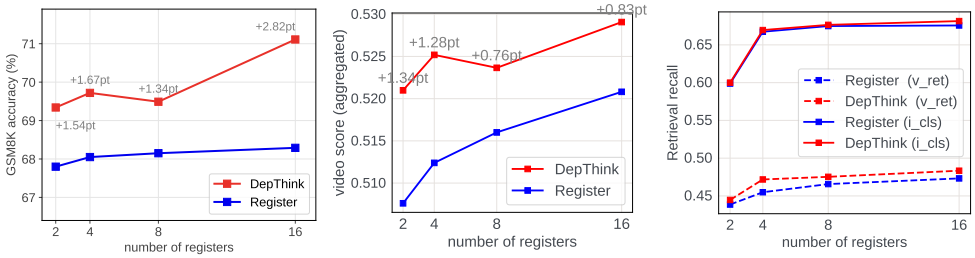
Table 4: Results for retrieval tasks. We denote the text, image, and video as t, i, and v, respectively, to show various "*query→target*" retrieval types. Table 8 reports the mean and standard deviation results of multiple runs for LoopLM and DepThink.

CIRR ($i+t \rightarrow i+t$)	
Method	Recall
CLIP (Baldrati et al., 2022)	42.8
BLIP (Liu et al., 2023)	44.8
VLM2Vec-7B (Jiang et al., 2024)	51.1
HINT (Zhang et al., 2026)	52.3
UME-R1-7B (Lan et al., 2025)	55.3
Compodiff (Gu et al., 2023)	64.5
LoopLM (2B)	60.9
Register (2B)	62.6
DepThink (2B)	**65.5**

MSCOCO T2I ($t \rightarrow i+t$)	
Method	Recall
CLIP (Radford et al., 2021)	37.8
CoCa (Yu et al., 2022)	51.2
BEIT-3 (Wang et al., 2023a)	66.0
Eval-CLIP (Sun et al., 2024)	73.6
VLM2Vec-7B (Jiang et al., 2024)	77.2
UME-R1-7B (Lan et al., 2025)	78.3
LoopLM (2B)	78.9
Register (2B)	78.6
DepThink (2B)	**79.3**

MSVD ($t \rightarrow v+t$)	
Method	Recall
CLIP4Clip (Luo et al., 2021)	45.2
VLM2Vec-7B (Jiang et al., 2024)	46.7
X-CLIP-B16 (Ma et al., 2022)	50.4
UMT (Li et al., 2023b)	58.2
UME-R1-7B (Lan et al., 2025)	60.8
InternVideo2 (Wang et al., 2024)	61.4
LoopLM (2B)	73.6
Register (2B)	71.5
DepThink (2B)	**77.2**

QVHighlights ($v+t \rightarrow v+t$)	
Method	Recall
UMT (Li et al., 2023b)	57.7
UME-R1-7B (Lan et al., 2025)	54.9
CLIP (Baldrati et al., 2022)	61.0
M-DETR (Lei et al., 2021)	60.2
TaskWeave (Yang et al., 2024)	64.3
UVCOM (Xiao et al., 2024)	64.5
LoopLM (2B)	**68.3**
Register (2B)	65.1
DepThink (2B)	67.9



(a) Math reasoning on GSM8K. (b) Video retrieval (c) Video RET. vs Image CLS.

Figure 2: Analysis of the number of registers for math reasoning and multimodal retrieval.

Knowledge question answering. Beyond math, we evaluate DepThink on knowledge-intensive open-domain and multi-hop question answering. Following prior work (Yue et al., 2026; Jin et al., 2025), we train models with $K = 16$ on the merged NQ (Kwiatkowski et al., 2019) and HotpotQA (Yang et al., 2018) training splits and evaluate on their test splits, in three settings: **Direct**, answering from the question alone; **Think**, generating a reasoning trace before the answer; and **Retrieve**, answering with the top-10 retrieved documents as context. Only HotpotQA provides the metadata needed for Think and Retrieve. As shown in Table 3, DepThink performs best in the Think and Retrieve settings and remains competitive in Direct, while LoopLM falls behind, especially in Retrieve.

Multimodal retrieval. Table 4 reports the performance of DepThink alongside recent baselines. We primarily compare DepThink against LoopLM and Register (both at 2B scale), and additionally include recent state-of-the-art methods on each dataset for reference. DepThink achieves the best performance on three out of four tasks (CIRR, MSCOCO T2I, and MSVD) and remains competitive on QVHighlights. Compared to the LoopLM and Register baselines, the gains are especially pronounced on CIRR (+4.6/+2.9 over LoopLM/Register) and MSVD (+3.6/+5.7), validating our depth-linked residual connections. On the QVHighlights, DepThink is only slightly worse than LoopLM (-0.4). In Appendix E.1, we performed multiple runs for both the LoopLM and our DepThink models with mean and standard deviation reported, showing robust separation from LoopLM.

5.2 Ablations and Analysis.

Register count scaling. We conduct experiments by varying the number of registers K for both text generation and multimodal retrieval tasks in Fig. 2. For text generation, the results in Fig. 2a on GSM8K math reasoning task shows that DepThink's math reasoning capability could largely be improved when K increases, and it consistently performs better than Register model. For retrieval tasks, we compare DepThink with the Register baseline by reporting the recall score in Figure 2b on video retrieval task. We again observe the large performance gain over the Register baseline.

Easy vs Hard retrieval tasks. To test DepThink's benefit on task difficulty, we compare DepThink and Register on video retrieval (v_ret), which requires fine-grained alignment between spatio-temporal video content and language, and zero-shot image classification (i_cls), a much easier task. As shown in Figure 2c, DepThink outperforms Register at every register count, with an average relative gain of 2.33% on v_ret, nearly six times the 0.41% on i_cls. This suggests that DepThink helps most on harder tasks: when visual-textual alignment is already easy, depth-linked residuals add little, but on v_ret the extra capacity helps capture the fine-grained semantics that retrieval needs.

Table 5: Model scaling on math reasoning (GSM8K) and multimodal retrieval (other 4 datasets).

Setting	GSM8K	CIRR	MSCOCO T2I	MSVD	QVHighlights
2B, LoRA	71.11	65.5	79.3	77.2	67.9
2B, full fine-tune	72.20	67.9	79.4	78.4	71.1
8B, LoRA	**85.76**	70.0	80.1	**86.7**	**74.4**
30B-A3B MoE, LoRA	84.31	**73.3**	**81.3**	84.0	71.5

Table 6: Average wall time in seconds per step on 16 H100 GPUs.

K	Training			Inference		
	Register	DepThink	LoopLM	Register	DepThink	LoopLM
2	1.74	1.74	3.13	1.66	1.69	2.01
4	1.80	1.91	5.77	1.80	1.71	2.73
8	1.75	1.81	11.25	1.69	1.68	4.20
16	1.78	1.85	22.04	1.67	1.70	7.24

Model scaling. In Table 5, we scale the DepThink model architecture up to 8B and 30B-A3B with MoE, and also compare between LoRA and full model fine-tuning recipes. These results showed that 8B and 30B-A3B models achieved much higher performance than 2B models for both tasks, while their performance differs by datasets. We can also observe that full model fine-tuning consistently shows better results than LoRA training. In Appendix E.2, we additionally include the LoopLM and Register models into comparison, showing that DepThink achieved better performance than LoopLM and Register models across different model scales..

Latency study. In Table 6, we benchmark DepThink, Register, and LoopLM models using 16 H100 GPUs, FSDP2, and bf16 precision. The per-GPU batch size is 32, the maximum sequence length is 2048, and a fixed seed provides the same batch stream. Training latency is averaged over 150 steps after 50 warm-up steps. Inference latency is averaged over 200 evaluation steps. DepThink and Register remains almost constant latency as K increases, showing that depth-linking itself adds little wall-clock overhead. LoopLM grows approximately linearly because it executes the backbone sequentially. At $K = 16$, DepThink is $11.9\times$ faster per training step and $4.3\times$ faster in inference.

Layer-wise extension. Since DepThink's two operations, registers and learnable residual links, are not tied to a specific layer, a natural extension is to apply them to multiple layers. Table 7 studies this on LLaVA-CoT (Xu et al., 2025) text generation. Adding registers and residual links to all layers lowers both perplexity (PPL) and cross-entropy (CE), as expected: the layer-wise version adds learnable parameters and gives each layer its own parameters for gathering residual information, making the registers both position- and depth-aware. Applying DepThink only to the middle layers performs best, even better than applying it to all layers, while applying it only to the late layers performs worst, suggesting DepThink is most useful for early-to-middle feature abstraction. Appendix E.3 visualizes $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ across layers and register positions, showing that $\boldsymbol{\alpha}$ is highest in the middle-to-late layers.

Table 7: Layer-wise extension for DepThink (28 layers in total).

No.	Registers	Residuals	PPL	CE
(1)	0	0-27	2.1647	0.7316
(2)	0-27	0-27	**2.0483**	**0.6814**
(3)	0-7	0-7	2.0455	0.6810
(4)	8-19	8-19	**2.0175**	**0.6677**
(5)	20-27	20-27	2.1591	0.7336

Quality of generated reasoning traces. We use Gemini-2.5-Pro as an LLM judge to score the semantic consistency and coherence of generated reasoning traces on OK-VQA (Marino et al., 2019), using the dataset's high-quality CoT annotations as references. As shown in Table 10 in Appendix E.4, DepThink scores highest on both metrics, outperforming Register and LoopLM. Appendix D provides the prompt used for the LLM-as-judge evaluation, and Appendix F provides qualitative examples showing how the generated traces change as K increases.

6 Conclusion

We introduced DepThink, a non-autoregressive latent reasoning framework that matches the expressiveness of looped language models while keeping the efficiency of parallel inference. By linking register embeddings across depth with gated residual connections, DepThink performs iterative refinement within a single pre-filling pass, avoiding the K-step autoregressive latency of LoopLMs. Theoretically, we showed that DepThink attains greater directional reachability and a one-step descent advantage over standard Transformers with registers. Empirically, DepThink outperforms both looped and register-based models on most reasoning-intensive text generation and retrieval benchmarks. Our study focuses on conditional text generation and representation learning; extending DepThink to full language-model pre-training at scale is a promising direction (Zhu et al., 2025; Zeng et al., 2026).

AI USE STATEMENT

In this work, we used generative AI tools for experiment planning, basic coding, reporting results, and improving manuscripts' readability. Additionally, we used generative AI tools for grammar checking to improve the readability, format references, and presenting figures and tables in latex. We have reviewed all AI-assisted work such as the core coding, reporting numbers, hallucination checking for references, etc. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, et al. Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, 35:23716–23736, 2022.

Thomas Bachlechner, Bodhisattwa Prasad Majumder, Huanru Henry Mao, Garrison W Cottrell, and Julian McAuley. Rezero is all you need: Fast convergence at large depth. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, 2021.

Alberto Baldrati, Marco Bertini, Tiberio Uricchio, and Alberto Del Bimbo. Conditioned and composed image retrieval combining and partially fine-tuning clip-based features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4959–4968, 2022.

David Chen and William B Dolan. Collecting highly parallel data for paraphrase evaluation. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pp. 190–200, 2011.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Rei Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Róbert Csordás, Kazuki Irie, and Jürgen Schmidhuber. The neural data router: Adaptive control flow in transformers improves systematic generalization. *arXiv:2110.07732*, 2021.

Wenliang Dai, Junnan Li, Dongxu Li, Anthony Tiong, Junqi Zhao, Weisheng Wang, Boyang Li, Pascale N Fung, and Steven Hoi. Instructblip: Towards general-purpose vision-language models with instruction tuning. *Advances in neural information processing systems*, 36:49250–49267, 2023.

Timothée Darcet, Maxime Oquab, Julien Mairal, and Piotr Bojanowski. Vision transformers need registers. *arXiv preprint arXiv:2309.16588*, 2023.

Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, and Łukasz Kaiser. Universal transformers. *arXiv preprint arXiv:1807.03819*, 2018.

Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, and Łukasz Kaiser. Universal transformers. In *ICLR*, 2019.

Yuntian Deng, Kiran Prasad, Roland Fernandez, Paul Smolensky, Vishrav Chaudhary, and Stuart Shieber. Implicit chain of thought reasoning via knowledge distillation. *arXiv preprint arXiv:2311.01460*, 2023.

Yuntian Deng, Yejin Choi, and Stuart Shieber. From explicit cot to implicit cot: Learning to internalize cot step by step. *arXiv preprint arXiv:2405.14838*, 2024.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Ying Fan, Yilun Du, Kannan Ramchandran, and Kangwook Lee. Looped transformers for length generalization. *arXiv preprint arXiv:2409.15647*, 2024.

Jonas Geiping, Sean McLeish, Neel Jain, John Kirchenbauer, Siddharth Singh, Brian R Bartoldson, Bhavya Kailkhura, Abhinav Bhatele, and Tom Goldstein. Scaling up test-time compute with latent reasoning: A recurrent depth approach. *arXiv preprint arXiv:2502.05171*, 2025.

Angeliki Giannou, Shashank Rajput, Jy-Yong Sohn, Kangwook Lee, Jason D. Lee, and Dimitris Papailiopoulos. Looped transformers as programmable computers. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 11398–11442. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/giannou23a.html.

Sachin Goyal, Ziwei Ji, Ankit Singh Rawat, Aditya Krishna Menon, Sanjiv Kumar, and Vaishnavh Nagarajan. Think before you speak: Training language models with pause tokens. *URL https://arxiv.org/abs/2310.02226*, 2024.

Geonmo Gu, Sanghyuk Chun, Wonjae Kim, HeeJae Jun, Yoohoon Kang, and Sangdoo Yun. Compodiff: Versatile composed image retrieval with latent diffusion. *arXiv preprint arXiv:2303.11916*, 2023.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

Shibo Hao, Sainbayar Sukhbaatar, DiJia Su, Xian Li, Zhiting Hu, Jason Weston, and Yuandong Tian. Training large language models to reason in a continuous latent space. *arXiv preprint arXiv:2412.06769*, 2024.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the math dataset. *NeurIPS*, 2021.

Ziyan Jiang, Rui Meng, Xinyi Yang, Semih Yavuz, Yingbo Zhou, and Wenhu Chen. Vlm2vec: Training vision-language models for massive multimodal embedding tasks. *arXiv preprint arXiv:2410.05160*, 2024.

Bowen Jin, Hansi Zeng, Zhenrui Yue, Jinsung Yoon, Sercan Arik, Dong Wang, Hamed Zamani, and Jiawei Han. Search-r1: Training llms to reason and leverage search engines with reinforcement learning. *arXiv preprint arXiv:2503.09516*, 2025.

Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, Kristina Toutanova, Llion Jones, Matthew Kelcey, Ming-Wei Chang, Andrew M. Dai, Jakob Uszkoreit, Quoc Le, and Slav Petrov. Natural questions: A benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, 7:452–466, 2019. doi: 10.1162/tacl_a_00276. URL https://aclanthology.org/Q19-1026/.

Zhibin Lan, Liqiang Niu, Fandong Meng, Jie Zhou, and Jinsong Su. Ume-r1: Exploring reasoning-driven generative multimodal embeddings. *arXiv preprint arXiv:2511.00405*, 2025.

Jie Lei, Tamara L Berg, and Mohit Bansal. Detecting moments and highlights in videos via natural language queries. *Advances in Neural Information Processing Systems*, 34:11846–11858, 2021.

Brian Lester, Rami Al-Rfou, and Noah Constant. The power of scale for parameter-efficient prompt tuning. In *EMNLP*, 2021.

Boxun Li, Yadong Li, Zhiyuan Li, Congyi Liu, Weilin Liu, Guowei Niu, Zheyue Tan, Haiyang Xu, Zhuyu Yao, Tao Yuan, et al. Megrez2 technical report. *arXiv preprint arXiv:2507.17728*, 2025.

Jiachun Li, Pengfei Cao, Zhuoran Jin, Yubo Chen, Jiexin Xu, Huaijun Li, Xiaojian Jiang, Kang Liu, and Jun Zhao. Fixing the broken compass: Diagnosing and improving inference-time reward modeling. In *International Conference on Learning Representations*, volume 2026, pp. 102566–102604, 2026.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, pp. 19730–19742. PMLR, 2023a.

Kunchang Li, Yali Wang, Yizhuo Li, Yi Wang, Yinan He, Limin Wang, and Yu Qiao. Unmasked teacher: Towards training-efficient video foundation models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 19948–19960, 2023b.

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *ACL*, 2021.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pp. 740–755. Springer, 2014.

Zheyuan Liu, Cristian Rodriguez-Opazo, Damien Teney, and Stephen Gould. Image retrieval on real-life images with pre-trained vision-and-language models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2125–2134, October 2021.

Zheyuan Liu, Weixuan Sun, Damien Teney, and Stephen Gould. Candidate set re-ranking for composed image retrieval with dual multi-modal encoder. *arXiv preprint arXiv:2305.16304*, 2023.

Huaishao Luo, Lei Ji, Ming Zhong, Yang Chen, Wen Lei, Nan Duan, and Tianrui Li. Clip4clip: An empirical study of clip for end to end video clip retrieval. *arXiv preprint arXiv:2104.08860*, 2021.

Yiwei Ma, Guohai Xu, Xiaoshuai Sun, Ming Yan, Ji Zhang, and Rongrong Ji. X-clip: End-to-end multi-grained contrastive learning for video-text retrieval. In *Proceedings of the 30th ACM international conference on multimedia*, pp. 638–647, 2022.

Kenneth Marino, Mohammad Rastegari, Ali Farhadi, and Roozbeh Mottaghi. Ok-vqa: A visual question answering benchmark requiring external knowledge. In *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, pp. 3195–3204, 2019.

Boris Mityagin. The zero set of a real analytic function. *arXiv preprint arXiv:1512.07276*, 2015.

Maxwell Nye, Anders Johan Andreassen, Guy Gur-Ari, Henryk Michalewski, Jacob Austin, David Bieber, David Dohan, Aitor Lewkowycz, Maarten Bosma, David Luan, Charles Sutton, and Augustus Odena. Show your work: Scratchpads for intermediate computation with language models, 2021.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Jacob Pfau, William Merrill, and Samuel R Bowman. Let's think dot by dot: Hidden computation in transformer language models. *arXiv preprint arXiv:2404.15758*, 2024.

Hayden Prairie, Zachary Novack, Taylor Berg-Kirkpatrick, and Daniel Y Fu. Parcae: Scaling laws for stable looped language models. *arXiv preprint arXiv:2604.12946*, 2026.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Nikunj Saunshi, Nishanth Dikkala, Zhiyuan Li, Sanjiv Kumar, and Sashank J Reddi. Reasoning with latent thoughts: On the power of looped transformers. *arXiv preprint arXiv:2502.17416*, 2025.

Xinyuan Song, Keyu Wang, PengXiang Li, Lu Yin, and Shiwei Liu. Demystifying the roles of llm layers in retrieval, knowledge, and reasoning. In *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 16792–16796. IEEE, 2026.

Rupesh Kumar Srivastava, Klaus Greff, and J"urgen Schmidhuber. Highway networks. *arXiv preprint arXiv:1505.00387*, 2015.

DiJia Su, Hanlin Zhu, Yingchen Xu, Jiantao Jiao, Yuandong Tian, and Qinqing Zheng. Token assorted: Mixing latent and text tokens for improved language model reasoning. *arXiv preprint arXiv:2502.03275*, 2025.

Quan Sun, Jinsheng Wang, Qiying Yu, Yufeng Cui, Fan Zhang, Xiaosong Zhang, and Xinlong Wang. Eva-clip-18b: Scaling clip to 18 billion parameters. *arXiv preprint arXiv:2402.04252*, 2024.

Hugo Touvron, Matthieu Cord, Alexandre Sablayrolles, Gabriel Synnaeve, and Hervé Jégou. Going deeper with image transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

Wenhui Wang, Hangbo Bao, Li Dong, Johan Bjorck, Zhiliang Peng, Qiang Liu, Kriti Aggarwal, Owais Khan Mohammed, Saksham Singhal, Subhojit Som, et al. Image as a foreign language: Beit pretraining for vision and vision-language tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19175–19186, 2023a.

Xinyi Wang, Lucas Caccia, Oleksiy Ostapenko, Xingdi Yuan, William Yang Wang, and Alessandro Sordoni. Guiding language model reasoning with planning tokens. *arXiv preprint arXiv:2310.05707*, 2023b.

Yi Wang, Kunchang Li, Xinhao Li, Jiashuo Yu, Yinan He, Guo Chen, Baoqi Pei, Rongkun Zheng, Zun Wang, Yansong Shi, et al. Internvideo2: Scaling foundation models for multimodal video understanding. In *European conference on computer vision*, pp. 396–416. Springer, 2024.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V. Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 24824–24837, 2022.

Bohong Wu, Mengzhao Chen, Xiang Luo, Shen Yan, Qifan Yu, Fan Xia, Tianqi Zhang, Hongrui Zhan, Zheng Zhong, Xun Zhou, et al. Parallel loop transformer for efficient test-time computation scaling. *arXiv preprint arXiv:2510.24824*, 2025.

Yicheng Xiao, Zhuoyan Luo, Yong Liu, Yue Ma, Hengwei Bian, Yatai Ji, Yujiu Yang, and Xiu Li. Bridging the gap: A unified video comprehension framework for moment retrieval and highlight detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 18709–18719, 2024.

Guowei Xu, Peng Jin, Ziang Wu, Hao Li, Yibing Song, Lichao Sun, and Li Yuan. Llava-cot: Let vision language models reason step-by-step. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2087–2098, 2025.

Kevin Xu and Issei Sato. On expressive power of looped transformers: Theoretical analysis and enhancement via timestep encoding. *arXiv preprint arXiv:2410.01405*, 2024.

Jin Yang, Ping Wei, Huan Li, and Ziyang Ren. Task-driven exploration: Decoupling and inter-task feedback for joint moment retrieval and highlight detection. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pp. 18308–18318, 2024.

Liu Yang, Kangwook Lee, Robert D Nowak, and Dimitris Papailiopoulos. Looped transformers are better at learning learning algorithms. In *The Twelfth International Conference on Learning Representations*, 2023.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. HotpotQA: A dataset for diverse, explainable multi-hop question answering. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.

Jiacheng Ye, Shansan Gong, Liheng Chen, Lin Zheng, Jiahui Gao, Han Shi, Chuan Wu, Xin Jiang, Zhenguo Li, Wei Bi, et al. Diffusion of thought: Chain-of-thought reasoning in diffusion language models. *Advances in Neural Information Processing Systems*, 37:105345–105374, 2024.

Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. Coca: Contrastive captioners are image-text foundation models. *arXiv preprint arXiv:2205.01917*, 2022.

Zhenrui Yue, Bowen Jin, Huimin Zeng, Honglei Zhuang, Zhen Qin, Jinsung Yoon, Lanyu Shang, Jiawei Han, and Dong Wang. Hybrid latent reasoning via reinforcement learning. *Advances in Neural Information Processing Systems*, 38:5501–5530, 2026.

Abbas Zeitoun, Lucas Torroba-Hennigen, and Yoon Kim. Hyperloop transformers. *arXiv preprint arXiv:2604.21254*, 2026.

Boyi Zeng, Shixiang Song, Siyuan Huang, Yixuan Wang, He Li, Ziwei He, Xinbing Wang, Zhouhan Lin, et al. Ponderlm: Pretraining language models to ponder in continuous space. In *The Fourteenth International Conference on Learning Representations*, 2026.

Mingyu Zhang, Zixu Li, Zhiwei Chen, Zhiheng Fu, Xiaowei Zhu, Jiajia Nie, Yinwei Wei, and Yupeng Hu. Hint: Composed image retrieval with dual-path compositional contextualized network. In *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 13002–13006. IEEE, 2026.

Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Conditional prompt learning for vision-language models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16816–16825, 2022a.

Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International journal of computer vision*, 130(9):2337–2348, 2022b.

Rui-Jie Zhu, Zixuan Wang, Kai Hua, Tianyu Zhang, Ziniu Li, Haoran Que, Boyi Wei, Zixin Wen, Fan Yin, He Xing, et al. Scaling latent reasoning via looped language models. *arXiv preprint arXiv:2510.25741*, 2025.

A Proof of Proposition 1

Proof. By fixing the IIR parameters to $\boldsymbol{\alpha} = \mathbf{1}$ and $\boldsymbol{\beta} = \mathbf{0}$, the recurrence in equation 3 simplifies to $\mathbf{z}_k = 0 \cdot \mathbf{z}_{k-1} + 1 \cdot \mathbf{e}_k = \mathbf{e}_k$ for all k. Thus, $\widehat{\mathbf{z}}_k = \widehat{\mathbf{e}}_k$ wherever $\|\mathbf{e}_k\|_2 > 0$, ensuring the heads of DepThink and Register yield identical multi-vector representations. This point-wise equivalence establishes the set inclusion $\mathcal{H}_\lambda \subseteq \mathcal{H}_{\text{dep}}$. The inequality regarding the optimal risks then follows directly from the fundamental property of infima: $\inf_{h \in \mathcal{H}_1} f(h) \geq \inf_{h \in \mathcal{H}_2} f(h)$ whenever $\mathcal{H}_1 \subseteq \mathcal{H}_2$. □

B Proof of Theorem 2

Proof. ($\subseteq$). By the closed-form expression in equation 4, $\mathbf{z}_k(x; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{k} M_{ki}(\boldsymbol{\alpha}, \boldsymbol{\beta})\, \mathbf{e}_i(x) \in \mathbb{V}_k(x)$. Therefore, its normalized counterpart satisfies $\widehat{\mathbf{z}}_k = \mathbf{z}_k / \|\mathbf{z}_k\|_2 \in \mathbb{V}_k(x) \cap \mathbb{S}^{D-1}$.

($\supseteq$). Let $\mathbf{u} \in \mathbb{V}_k(x) \cap \mathbb{S}^{D-1}$. By the definition of $\mathbb{V}_k$, there exist coefficients $c_1, \ldots, c_k \in \mathbb{R}$ such that $\mathbf{u} = \sum_{i=1}^{k} c_i\, \mathbf{e}_i(x)$. We construct the specific IIR coefficients:

$$\boldsymbol{\alpha}^* \triangleq [c_1, c_2, \ldots, c_k, 0, \ldots, 0] \in \mathbb{R}^K,$$
$$\boldsymbol{\beta}^* \triangleq [0, 1, 1, \ldots, 1] \in \mathbb{R}^K.$$

Both rigorously lie within their respective parameter spaces. For $i \leq k$, the product term simplifies to $\prod_{j=i+1}^{k} \beta_j^* = 1$, yielding $M_{ki}(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = c_i$. Substituting these analytical weights into the closed form gives:

$$\mathbf{z}_k(x; \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = \sum_{i=1}^{k} c_i\, \mathbf{e}_i(x) = \mathbf{u}.$$

Because $\|\mathbf{u}\|_2 = 1$, we obtain $\|\mathbf{z}_k\|_2 = 1 > 0$, ensuring $\widehat{\mathbf{z}}_k(x; \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = \mathbf{u}$. For the standard Register scheme, no IIR coefficients exist, and therefore, the vector $\widehat{\mathbf{e}}_k(x) = \mathbf{e}_k(x)/\|\mathbf{e}_k(x)\|_2$ is deterministically locked by the tuple $(\boldsymbol{\theta}, \mathbf{R}, x)$. □

C Extended Theoretical Analysis and Deferred Proofs for Section 4.3

In this section, we provide the detailed analytical derivations, auxiliary lemmas, and the formal proof for the optimization advantage established in Section 4.3.

C.1 Partial Derivatives at the Lift Point

We first compute the exact partial derivatives of the DepThink representation with respect to its recursive parameters, evaluated at the initial lift point.

Lemma 5 (Partial derivatives at the lift)**.** *For any $k \geq 2$ and $i \in \{1, \ldots, K\}$, the partial derivatives of the unnormalized DepThink representation $\mathbf{z}_k$ at the lift point $\boldsymbol{\eta}_0$ are given by*

$$\left.\frac{\partial \mathbf{z}_k}{\partial \alpha_i}\right|_{\boldsymbol{\eta}_0} = \begin{cases} \mathbf{e}_k(x), & i = k, \\ \mathbf{0}, & i \neq k, \end{cases}$$
$$\left.\frac{\partial \mathbf{z}_k}{\partial \beta_i}\right|_{\boldsymbol{\eta}_0} = \begin{cases} \mathbf{e}_{k-1}(x), & i = k, \\ \mathbf{0}, & i \neq k. \end{cases} \tag{16}$$

Furthermore, for the ℓ_2-normalized representation $\widehat{\mathbf{z}}_k$, we have

$$\left.\frac{\partial \widehat{\mathbf{z}}_k}{\partial \alpha_k}\right|_{\boldsymbol{\eta}_0} = \mathbf{0},$$
$$\left.\frac{\partial \widehat{\mathbf{z}}_k}{\partial \beta_k}\right|_{\boldsymbol{\eta}_0} = \frac{\boldsymbol{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k(x)}\, \mathbf{e}_{k-1}(x)}{\|\mathbf{e}_k(x)\|_2}, \tag{17}$$

where $\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k(x)}$ is the orthogonal projection operator that extracts the component orthogonal to $\widehat{\mathbf{e}}_k(x)$. Specifically,

$$\begin{aligned}\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k(x)}\,\mathbf{e}_{k-1}(x) = \mathbf{e}_{k-1}(x) \\ - \langle \widehat{\mathbf{e}}_k(x), \mathbf{e}_{k-1}(x)\rangle\, \widehat{\mathbf{e}}_k(x) \in \mathbb{R}^D.\end{aligned} \tag{18}$$

For the base case where $k = 1$, the derivatives evaluate to zero: $\partial \mathbf{z}_1/\partial\beta_1 = \mathbf{z}_0 = \mathbf{0}$, $\partial\widehat{\mathbf{z}}_1/\partial\beta_1 = \mathbf{0}$, and $\partial\widehat{\mathbf{z}}_1/\partial\alpha_1 = \mathbf{0}$.

Proof. By differentiating $M_{ki}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \alpha_i \prod_{j=i+1}^{k} \beta_j$ and applying the chain rule to the expansion $\mathbf{z}_k = \sum_i M_{ki}\,\mathbf{e}_i$, we obtain

$$\begin{aligned}\frac{\partial \mathbf{z}_k}{\partial \alpha_i} &= \Big(\prod_{j=i+1}^{k} \beta_j\Big)\mathbf{e}_i, \\ \frac{\partial \mathbf{z}_k}{\partial \beta_{i'}} &= \sum_{i<i'} \alpha_i \mathbf{e}_i \prod_{\substack{j=i+1 \\ j\neq i'}}^{k} \beta_j.\end{aligned} \tag{19}$$

At the lift point $\boldsymbol{\eta}_0$, $\boldsymbol{\beta} = \mathbf{0}$, and hence the product $\prod_{j\in\mathcal{J}} \beta_j$ is non-zero *if and only if* the index set $\mathcal{J}$ is empty. This condition restricts the α-derivative to the case where $i = k$, yielding $\partial\mathbf{z}_k/\partial\alpha_k|_{\boldsymbol{\eta}_0} = \mathbf{e}_k(x)$ and zero otherwise. Similarly, for the β-derivative, the index set $\{i+1, \ldots, k\}\setminus\{i'\}$ is empty *if and only if* $i = k-1$ and $i' = k$. Given $\alpha_{k-1} = 1$ at the lift point, this yields $\partial\mathbf{z}_k/\partial\beta_k|_{\boldsymbol{\eta}_0} = \mathbf{e}_{k-1}(x)$ and zero otherwise, which establishes equation 16.

Next, the ℓ_2-normalization function $\boldsymbol{\phi}(\mathbf{u}) = \mathbf{u}/\|\mathbf{u}\|_2$ has the Jacobian matrix $D\boldsymbol{\phi}(\mathbf{u}) = \frac{1}{\|\mathbf{u}\|_2}\mathbf{\Pi}^{\perp}_{\mathbf{u}/\|\mathbf{u}\|_2}$. Applying the multivariable chain rule for any parameter $\xi \in \{\alpha_i, \beta_i\}_{i=1}^{k}$, we have

$$\begin{aligned}\frac{\partial\widehat{\mathbf{z}}_k}{\partial\xi}\bigg|_{\boldsymbol{\eta}_0} &= D\boldsymbol{\phi}(\mathbf{z}_k)\bigg|_{\boldsymbol{\eta}_0}\left[\frac{\partial\mathbf{z}_k}{\partial\xi}\bigg|_{\boldsymbol{\eta}_0}\right] \\ &= \frac{1}{\|\mathbf{e}_k\|_2}\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k}\left[\frac{\partial\mathbf{z}_k}{\partial\xi}\bigg|_{\boldsymbol{\eta}_0}\right].\end{aligned} \tag{20}$$

by employing the fact that $\mathbf{z}_k|_{\boldsymbol{\eta}_0} = \mathbf{e}_k$ and $\mathbf{z}_k/\|\mathbf{z}_k\|_2|_{\boldsymbol{\eta}_0} = \widehat{\mathbf{e}}_k$.

For $\xi = \alpha_k$, we substitute $\partial\mathbf{z}_k/\partial\alpha_k|_{\boldsymbol{\eta}_0} = \mathbf{e}_k(x)$. Since $\mathbf{e}_k$ is parallel to $\widehat{\mathbf{e}}_k$, the orthogonal projector returns zero, *i.e.,* , $\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k}\mathbf{e}_k = \mathbf{0}$. For $\xi = \beta_k$, we substitute $\partial\mathbf{z}_k/\partial\beta_k|_{\boldsymbol{\eta}_0} = \mathbf{e}_{k-1}(x)$, which gives the expression in equation 17. The case where $k = 1$ trivially vanishes because $\mathbf{z}_0 \triangleq \mathbf{0}$, completing the proof. □

By using Lemma 5, we can now compute the closed-form expression for the gradient of the empirical InfoNCE risk with respect to β_k.

Theorem 6 (Closed-form Empirical InfoNCE Risk Gradient at the Lift)**.** *Let $p_{ij} \triangleq \exp(S^{\mathrm{dep}}(x_i, y_j)/\tau)/\sum_{l=1}^{N}\exp(S^{\mathrm{dep}}(x_i, y_l)/\tau)$ denote the InfoNCE softmax probability associated with the DepThink score. Evaluated at the lift point $\boldsymbol{\eta}_0$, these probabilities match the Register baseline ($p_{ij}|_{\boldsymbol{\eta}_0} = p^{\lambda}_{ij}$). For any depth $k \geq 2$, the partial derivative of the empirical risk with respect to β_k is given by:*

$$\begin{aligned}\tau\frac{\partial\hat{L}_N(\mathbf{\Phi}^{\mathrm{dep}})}{\partial\beta_k}\bigg|_{\boldsymbol{\eta}_0} = \frac{1}{NK}\sum_{i,j=1}^{N}(p_{ij} - \delta_{ij})\Bigg[\left\langle\frac{\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k(x_i)}\mathbf{e}_{k-1}(x_i)}{\|\mathbf{e}_k(x_i)\|_2}, \widehat{\mathbf{e}}_k(y_j)\right\rangle \\ + \left\langle\widehat{\mathbf{e}}_k(x_i), \frac{\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k(y_j)}\mathbf{e}_{k-1}(y_j)}{\|\mathbf{e}_k(y_j)\|_2}\right\rangle\Bigg],\end{aligned} \tag{21}$$

where δ_{ij} is the Kronecker delta.

Proof. For a fixed query index i, the per-query empirical InfoNCE loss is defined as

$$q_i \triangleq -\log p_{ii} = -\frac{S^{\mathrm{dep}}(x_i, y_i)}{\tau} + \log \sum_{l=1}^{N} \exp\big(S^{\mathrm{dep}}(x_i, y_l)/\tau\big).$$

Using the standard properties of the softmax cross-entropy loss, the partial derivative of q_i with respect to a single score $S^{\mathrm{dep}}(x_i, y_j)$ is

$$\begin{aligned} \frac{\partial q_i}{\partial S^{\mathrm{dep}}(x_i, y_j)} &= -\frac{\delta_{ij}}{\tau} + \frac{1}{\tau} \frac{\exp(S^{\mathrm{dep}}(x_i, y_j)/\tau)}{\sum_l \exp(S^{\mathrm{dep}}(x_i, y_l)/\tau)} \\ &= \frac{1}{\tau}(p_{ij} - \delta_{ij}). \end{aligned} \tag{22}$$

Notice that $q_{i'}$ depends only on the scores $\{S^{\mathrm{dep}}(x_{i'}, y_l)\}_{l=1}^{N}$, so $\partial q_{i'}/\partial S^{\mathrm{dep}}(x_i, y_j) = 0$ for all $i' \neq i$. Since the overall empirical risk is $\hat{L}_N(\mathbf{\Phi}^{\mathrm{dep}}) = N^{-1} \sum_{i'} q_{i'}$, we have

$$\begin{aligned} \frac{\partial \hat{L}_N(\mathbf{\Phi}^{\mathrm{dep}})}{\partial S^{\mathrm{dep}}(x_i, y_j)} &= \frac{1}{N} \sum_{i'=1}^{N} \frac{\partial q_{i'}}{\partial S^{\mathrm{dep}}(x_i, y_j)} \\ &= \frac{1}{N} \frac{\partial q_i}{\partial S^{\mathrm{dep}}(x_i, y_j)} = \frac{p_{ij} - \delta_{ij}}{N\tau}. \end{aligned} \tag{23}$$

By applying the product rule to the score function, the derivative with respect to β_k is

$$\begin{aligned} \frac{\partial S^{\mathrm{dep}}(x_i, y_j)}{\partial \beta_k} = \frac{1}{K} \sum_{k'=1}^{K} \Bigg(&\Big\langle \frac{\partial \widehat{\mathbf{z}}_{k'}(x_i)}{\partial \beta_k}, \widehat{\mathbf{z}}_{k'}(y_j) \Big\rangle \\ &+ \Big\langle \widehat{\mathbf{z}}_{k'}(x_i), \frac{\partial \widehat{\mathbf{z}}_{k'}(y_j)}{\partial \beta_k} \Big\rangle \Bigg). \end{aligned} \tag{24}$$

We evaluate this expression at the lift point $\boldsymbol{\eta}_0$. According to equation 17 in Lemma 5, the derivatives $\partial \widehat{\mathbf{z}}_{k'}(x)/\partial \beta_k$ are sparse:

$$\left.\frac{\partial \widehat{\mathbf{z}}_{k'}(x)}{\partial \beta_k}\right|_{\boldsymbol{\eta}_0} = \begin{cases} \dfrac{\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k(x)} \mathbf{e}_{k-1}(x)}{\|\mathbf{e}_k(x)\|_2}, & k' = k, \\ \mathbf{0}, & k' \neq k, \end{cases} \tag{25}$$

and similarly for any target y. Substituting these results into equation 24 means only the term where $k' = k$ remains, that is,

$$\begin{aligned} \left.\frac{\partial S^{\mathrm{dep}}(x_i, y_j)}{\partial \beta_k}\right|_{\boldsymbol{\eta}_0} = \frac{1}{K} \Bigg[&\Big\langle \frac{\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k(x_i)} \mathbf{e}_{k-1}(x_i)}{\|\mathbf{e}_k(x_i)\|_2}, \widehat{\mathbf{e}}_k(y_j) \Big\rangle \\ &+ \Big\langle \widehat{\mathbf{e}}_k(x_i), \frac{\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k(y_j)} \mathbf{e}_{k-1}(y_j)}{\|\mathbf{e}_k(y_j)\|_2} \Big\rangle \Bigg]. \end{aligned} \tag{26}$$

Finally, by applying the multivariable chain rule with equation 23 and equation 26, and multiplying by τ, we obtain

$$\begin{aligned} \tau \left.\frac{\partial \hat{L}_N(\mathbf{\Phi}^{\mathrm{dep}})}{\partial \beta_k}\right|_{\boldsymbol{\eta}_0} &= \tau \sum_{i,j=1}^{N} \frac{\partial \hat{L}_N(\mathbf{\Phi}^{\mathrm{dep}})}{\partial S^{\mathrm{dep}}(x_i, y_j)} \left.\frac{\partial S^{\mathrm{dep}}(x_i, y_j)}{\partial \beta_k}\right|_{\boldsymbol{\eta}_0} \\ &= \frac{1}{NK} \sum_{i,j=1}^{N} (p_{ij} - \delta_{ij}) \cdot \Bigg[\Big\langle \frac{\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k(x_i)} \mathbf{e}_{k-1}(x_i)}{\|\mathbf{e}_k(x_i)\|_2}, \widehat{\mathbf{e}}_k(y_j) \Big\rangle \\ &\quad + \Big\langle \widehat{\mathbf{e}}_k(x_i), \frac{\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{e}}_k(y_j)} \mathbf{e}_{k-1}(y_j)}{\|\mathbf{e}_k(y_j)\|_2} \Big\rangle \Bigg], \end{aligned} \tag{27}$$

which completes the proof. □

C.2 Gradient Identities and Descent Gap Expansion

To analyze the gradient descent dynamics, we first make a standard local smoothness assumption required for bounding the descent trajectory.

Assumption 1 (Local smoothness near the lift)**.** For any given backbone parameters $(\boldsymbol{\theta}, \mathbf{R}) \in \boldsymbol{\Theta} \times \mathbb{R}^{K\times D}$, there exists an open neighborhood U_0 around the lift point $\boldsymbol{\eta}_0 \triangleq (\boldsymbol{\theta}, \mathbf{R}, \mathbf{1}, \mathbf{0})$ where the expected InfoNCE risk $L_{\mathcal{P}}$ is twice continuously differentiable and its gradient is Lipschitz continuous with constant $L_{\mathrm{loc}} > 0$, that is,

$$\|\nabla L_{\mathcal{P}}(\boldsymbol{\eta}) - \nabla L_{\mathcal{P}}(\boldsymbol{\eta}')\|_2 \le L_{\mathrm{loc}}\|\boldsymbol{\eta} - \boldsymbol{\eta}'\|_2, \quad \text{for all } \boldsymbol{\eta}, \boldsymbol{\eta}' \in U_0. \tag{28}$$

Assumption 1 is a standard local non-convex regularity assumption which is widely adopted in the theoretical optimization literature.

Before comparing the gradient steps, we first provide a lemma that reveals the relationships between the gradient of DepThink and that of Register at the exact point of lifting.

Lemma 7 (Lift Gradient Identity)**.** *At the lift point $\boldsymbol{\eta}_0(\boldsymbol{\theta}, \mathbf{R})$, the following identities hold*

(i) $L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}}_{\boldsymbol{\eta}_0}) = L_{\mathcal{P}}(\boldsymbol{\Phi}^{\lambda}_{(\boldsymbol{\theta},\mathbf{R})})$.
(ii) $\nabla_{(\boldsymbol{\theta},\mathbf{R})} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} = \nabla_{(\boldsymbol{\theta},\mathbf{R})} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\lambda})|_{(\boldsymbol{\theta},\mathbf{R})}$.
(iii) $\nabla_{\boldsymbol{\alpha}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} = \mathbf{0}$.
(iv) $\|\nabla L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0}\|^2 = \|\nabla_{(\boldsymbol{\theta},\mathbf{R})} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\lambda})|_{(\boldsymbol{\theta},\mathbf{R})}\|^2 + \|\nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0}\|^2$.

Proof. Identity (i) is a direct restatement of Proposition 1.

For (ii), note that restricting the function $L^{\mathrm{dep}}_{\mathcal{P}}(\boldsymbol{\theta}, \mathbf{R}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ to the condition where $\boldsymbol{\alpha} = \mathbf{1}$ and $\boldsymbol{\beta} = \mathbf{0}$ perfectly recovers $L^{\lambda}_{\mathcal{P}}(\boldsymbol{\theta}, \mathbf{R})$. Since the partial derivatives with respect to $(\boldsymbol{\theta}, \mathbf{R})$ are taken strictly along this condition, they are equal.

To prove (iii), we consider a neighborhood where $\boldsymbol{\alpha} > \mathbf{0}$ and $\boldsymbol{\beta} = \mathbf{0}$. In this region, the latent thought $\mathbf{z}_k$ simplifies to $\mathbf{z}_k = \alpha_k \mathbf{e}_k$. Because the ℓ_2-normalization is scale-invariant, we have $\widehat{\mathbf{z}}_k = (\alpha_k \mathbf{e}_k)/\|\alpha_k \mathbf{e}_k\|_2 = \widehat{\mathbf{e}}_k$ for any $\alpha_k > 0$. Thus, the normalized $\widehat{\mathbf{z}}_k$ and the expected risk $L^{\mathrm{dep}}_{\mathcal{P}}$ are completely independent of $\boldsymbol{\alpha}$ near the lift point. Therefore, the gradient with respect to $\boldsymbol{\alpha}$ is zero, that is, $\nabla_{\boldsymbol{\alpha}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} = \mathbf{0}$.

Finally, (iv) holds because the parameter groups $(\boldsymbol{\theta}, \mathbf{R}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ are independent, meaning the squared ℓ_2-norm of the full gradient decomposes into the sum of the squared norms of its parts, that is, $\|\nabla L^{\mathrm{dep}}_{\mathcal{P}}\|^2 = \|\nabla_{(\boldsymbol{\theta},\mathbf{R})} L^{\mathrm{dep}}_{\mathcal{P}}\|^2 + \|\nabla_{\boldsymbol{\alpha}} L^{\mathrm{dep}}_{\mathcal{P}}\|^2 + \|\nabla_{\boldsymbol{\beta}} L^{\mathrm{dep}}_{\mathcal{P}}\|^2$. Substituting (ii) and (iii) directly yields the result. □

We now develop the closed-form expression for the descent gap after a single gradient step.

Lemma 8 (Descent Gap Expansion)**.** *Suppose Assumption 1 holds, meaning the expected InfoNCE risk has an L_{loc}-Lipschitz continuous gradient within a local neighborhood U_0 of radius r_{U_0}. Let $(\boldsymbol{\theta}', \mathbf{R}') \triangleq (\boldsymbol{\theta}, \mathbf{R}) - \mu \mathbf{g}^{\lambda}$ and $\boldsymbol{\eta}' \triangleq \boldsymbol{\eta}_0 - \mu \mathbf{g}^{\mathrm{dep}}$ be the parameters updated by a single gradient descent step with step size $\mu > 0$, where $\mathbf{g}^{\lambda} \triangleq \nabla L_{\mathcal{P}}(\boldsymbol{\Phi}^{\lambda})|_{(\boldsymbol{\theta},\mathbf{R})}$ and $\mathbf{g}^{\mathrm{dep}} \triangleq \nabla L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0}$. Define the maximum admissible step size ensuring updates remain within U_0 and satisfy descent conditions as*

$$\mu_{\max} \triangleq \min\left(\frac{1}{L_{\mathrm{loc}}}, \frac{r_{U_0}}{\|\mathbf{g}^{\mathrm{dep}}\|}, \frac{r_{U_0}}{\|\mathbf{g}^{\lambda}\|}\right). \tag{29}$$

Define the one-step descent gap between the expected risk of Register and that of DepThink as

$$\Delta(\mu) \;\triangleq\; L_{\mathcal{P}}\big(\boldsymbol{\Phi}^{\lambda}_{(\boldsymbol{\theta}',\mathbf{R}')}\big) - L_{\mathcal{P}}\big(\boldsymbol{\Phi}^{\mathrm{dep}}_{\boldsymbol{\eta}'}\big). \tag{30}$$

For any step size bounded by $\mu \le \mu_{\max}$, the gap has the exact expansion

$$\Delta(\mu) = \mu\big\|\nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0}\big\|^2 + O(\mu^2). \tag{31}$$

Specifically, the remainder term $O(\mu^2)$ is bounded by $|O(\mu^2)| \le \frac{\mu^2}{2} L_{\mathrm{loc}}\big(\|\mathbf{g}^{\lambda}\|^2 + \|\mathbf{g}^{\mathrm{dep}}\|^2\big)$.

Proof. For DepThink, applying Taylor's theorem along the update direction yields

$$L_{\mathcal{P}}^{\mathrm{dep}}(\boldsymbol{\eta}_0 - \mu \mathbf{g}^{\mathrm{dep}}) = L_{\mathcal{P}}^{\mathrm{dep}}(\boldsymbol{\eta}_0) - \mu \|\mathbf{g}^{\mathrm{dep}}\|^2 + R^{\mathrm{dep}}, \tag{32}$$

where the remainder term is

$$R^{\mathrm{dep}} \triangleq \int_0^{\mu} \langle \nabla L_{\mathcal{P}}^{\mathrm{dep}}(\boldsymbol{\eta}_0 - t\mathbf{g}^{\mathrm{dep}}) - \nabla L_{\mathcal{P}}^{\mathrm{dep}}(\boldsymbol{\eta}_0), -\mathbf{g}^{\mathrm{dep}} \rangle \, dt. \tag{33}$$

By bounding the step size as $\mu \leq \mu_{\max}$, we guarantee that the update distance $\mu\|\mathbf{g}^{\mathrm{dep}}\| \leq r_{U_0}$, ensuring the trajectory strictly remains within the local neighborhood U_0 where Assumption 1 holds. Furthermore, because $\mu \leq 1/L_{\mathrm{loc}}$, we can tightly bound the remainder using the Cauchy-Schwarz inequality and the L_{loc}-Lipschitz gradient condition

$$\begin{aligned} |R^{\mathrm{dep}}| &\leq \int_0^{\mu} \left\| \nabla L_{\mathcal{P}}^{\mathrm{dep}}(\boldsymbol{\eta}_0 - t\mathbf{g}^{\mathrm{dep}}) - \nabla L_{\mathcal{P}}^{\mathrm{dep}}(\boldsymbol{\eta}_0) \right\|_2 \|\mathbf{g}^{\mathrm{dep}}\|_2 \, dt \\ &\leq \int_0^{\mu} L_{\mathrm{loc}} \, t \, \|\mathbf{g}^{\mathrm{dep}}\|_2^2 \, dt \\ &= \frac{\mu^2}{2} L_{\mathrm{loc}} \|\mathbf{g}^{\mathrm{dep}}\|^2. \end{aligned} \tag{34}$$

Similarly, the standard Register update has the following Taylor expansion

$$L_{\mathcal{P}}^{\lambda}((\boldsymbol{\theta}, \mathbf{R}) - \mu \mathbf{g}^{\lambda}) = L_{\mathcal{P}}^{\lambda}(\boldsymbol{\theta}, \mathbf{R}) - \mu \|\mathbf{g}^{\lambda}\|^2 + R^{\lambda}, \tag{35}$$

where the remainder is bounded by

$$|R^{\lambda}| \leq \frac{\mu^2}{2} L_{\mathrm{loc}} \|\mathbf{g}^{\lambda}\|^2. \tag{36}$$

By subtracting the DepThink expansion equation 32 from the Register expansion equation 35, and using the equality $L_{\mathcal{P}}^{\mathrm{dep}}(\boldsymbol{\eta}_0) = L_{\mathcal{P}}^{\lambda}(\boldsymbol{\theta}, \mathbf{R})$ from Lemma 7 (i), we have

$$\begin{aligned} \Delta(\mu) &= \mu \left(\|\mathbf{g}^{\mathrm{dep}}\|^2 - \|\mathbf{g}^{\lambda}\|^2 \right) + (R^{\lambda} - R^{\mathrm{dep}}) \\ &= \mu \left\| \nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}^{\mathrm{dep}}|_{\boldsymbol{\eta}_0} \right\|^2 + (R^{\lambda} - R^{\mathrm{dep}}), \end{aligned}$$

where the second line uses the gradient decomposition from Lemma 7 (iv). The bound on the total remainder follows directly from the triangle inequality

$$|R^{\lambda} - R^{\mathrm{dep}}| \leq |R^{\lambda}| + |R^{\mathrm{dep}}| \leq \frac{\mu^2}{2} L_{\mathrm{loc}} \left(\|\mathbf{g}^{\lambda}\|^2 + \|\mathbf{g}^{\mathrm{dep}}\|^2 \right). \tag{37}$$

□

C.3 Proof of Theorem ~~1~~ 4: Weak and Strict Optimization Advantage

We are now ready to prove the main theorem presented in Section ~~5.3~~ 4.3.

Restatement of Theorem 4 Suppose the expected InfoNCE risk satisfies the local smoothness Assumption 1 in Appendix C with Lipschitz constant $L_{\mathrm{loc}} > 0$, admitting a maximum valid step size $\mu_{\max} > 0$ that confines the update to the locally Lipschitz neighborhood. Define the critical step-size threshold

$$\bar{\mu} \triangleq \frac{2 \left\| \nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} \right\|^2}{L_{\mathrm{loc}} \left(\|\mathbf{g}^{\lambda}\|^2 + \|\mathbf{g}^{\mathrm{dep}}\|^2 \right)}. \tag{38}$$

Then, for any gradient descent step size $\mu \in \left(0, \min(\bar{\mu}, \mu_{\max})\right)$, the following properties hold

1. **Weak Optimization Advantage:** $\Delta(\mu) \geq 0$ whenever $\mu \in \left(0, \min(\bar{\mu}, \mu_{\max})\right)$. Furthermore, the exact equality $\Delta(\mu) = 0$ holds if and only if $\nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} = \mathbf{0}$, in which case the descent trajectories of DepThink and Register perfectly coincide.
2. **Strict Optimization Advantage:** If $\nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} \neq \mathbf{0}$, then $\Delta(\mu) > 0$ strictly for all $\mu \in \left(0, \min(\bar{\mu}, \mu_{\max})\right)$, guaranteeing that DepThink consistently achieves a strictly greater InfoNCE risk reduction than Register.

Proof. First, consider the case where the gradient with respect to $\boldsymbol{\beta}$ is zero, that is, $\nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} = \mathbf{0}$. From Lemma 7 (ii) and (iii), the full gradient $\mathbf{g}^{\mathrm{dep}}$ is simply $(\mathbf{g}^{\lambda}, \mathbf{0}, \mathbf{0})$. This means that after a gradient descent step, the DepThink parameters update to $\boldsymbol{\eta}' = (\boldsymbol{\theta}', \mathbf{R}', \mathbf{1}, \mathbf{0})$. Notice that this is exactly the same as applying the lift operation to the newly updated Register parameters. According to Proposition 1, the functions computed by both DepThink and Register are identical, so their expected risks are equal. Thus, $\Delta(\mu) = 0$ for all $\mu > 0$. This completes the proof of the Weak Optimization Advantage boundary condition, showing that the trajectories coincide when there is no useful gradient signal from the recursive parameters.

Next, consider the case where $\nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} \neq \mathbf{0}$. Provided $\mu \leq \mu_{\max}$, we can use the exact expansion and the remainder bound from Lemma 8 to establish a strict lower bound on the descent gap

$$\Delta(\mu) \;\geq\; \mu \big\| \nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} \big\|^2 \;-\; \frac{\mu^2}{2} L_{\mathrm{loc}} \big(\|\mathbf{g}^{\lambda}\|^2 + \|\mathbf{g}^{\mathrm{dep}}\|^2 \big). \tag{39}$$

For DepThink to strictly outperform Register in the single step (*i.e.,* $\Delta(\mu) > 0$), the linear gain must be strictly larger than the quadratic penalty

$$\mu \big\| \nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} \big\|^2 \;>\; \frac{\mu^2}{2} L_{\mathrm{loc}} \big(\|\mathbf{g}^{\lambda}\|^2 + \|\mathbf{g}^{\mathrm{dep}}\|^2 \big). \tag{40}$$

Since $\mu > 0$, we can divide by μ and rearrange the inequality to find the condition for μ

$$\mu < \frac{2 \big\| \nabla_{\boldsymbol{\beta}} L_{\mathcal{P}}(\boldsymbol{\Phi}^{\mathrm{dep}})|_{\boldsymbol{\eta}_0} \big\|^2}{L_{\mathrm{loc}} \left(\|\mathbf{g}^{\lambda}\|^2 + \|\mathbf{g}^{\mathrm{dep}}\|^2 \right)} = \bar{\mu}. \tag{41}$$

To simultaneously satisfy this advantage condition and ensure the theoretical validity of the Taylor expansion which strictly requires $\mu \leq \mu_{\max}$, the step size must be bounded by both thresholds. Therefore, whenever μ is in the interval $\big(0, \min(\bar{\mu}, \mu_{\max})\big)$, the Strict Optimization Advantage holds, completing the proof. □

D PROMPTS FOR LLMS AS JUDGES

LLM-as-a-Judge Evaluation Rubric

System Prompt:
You are an expert evaluator specializing in assessing the quality of Chain-of-Thought (COT) reasoning for visual understanding tasks. You will be given an image, a task description, a ground-truth reference answer, and a COT reasoning trace. Your job is to score the COT on two dimensions: Semantic Accuracy and Coherence.

1. Semantic Accuracy (1–5)
Does the COT correctly capture the essential meaning, facts, and logical relationships needed to solve the task? Does it arrive at (or support) the correct answer as given by the ground truth?

Score	Description
5	All key facts, entities, and relationships are correctly identified. The reasoning leads to the correct answer consistent with the ground truth. No factual errors or hallucinations.
4	Most key information is correct with only minor omissions. The final answer aligns with the ground truth. No significant factual errors.
3	Core meaning is partially captured but some important details are missing or slightly inaccurate. The answer may partially match the ground truth.
2	Multiple factual errors or significant omissions. The reasoning leads to an answer that deviates substantially from the ground truth.
1	Mostly incorrect, irrelevant, or hallucinated content. The conclusion contradicts the ground truth.

2. Coherence (1–5)
Is the COT logically structured, internally consistent, and easy to follow as a reasoning chain?

Score	Description
5	Clear logical flow from observation to conclusion. Each step follows naturally from the previous one. Well-organized and concise.
4	Generally logical with minor gaps or redundancies. The overall reasoning direction is clear.
3	Partially coherent but contains noticeable logical jumps, contradictions, or disorganized segments.
2	Fragmented reasoning with weak connections between steps. Hard to follow the logical chain.
1	Incoherent, self-contradictory, or completely disorganized. No discernible reasoning structure.

E ADDITIONAL EMPIRICAL RESULTS

E.1 REPEATED-RUN FOR MULTIMODEL RETRIEVAL

To quantify run-to-run variability, we repeated the controlled 2B/LoRA comparison three times with independent runs and otherwise identical settings. Table 8 reports the mean and sample standard deviation. The CIRR and MSVD gaps are large relative to the observed variability, while MSCOCO T2I and QVHighlights should be interpreted as ties. DepThink nevertheless has the highest four-task mean. On QVHighlights, its standard deviation is approximately eight times smaller than LoopLM's, suggesting greater stability in these runs; establishing this trend more broadly will require additional repetitions.

Table 8: Multi-run retrieval results under the identical 2B/LoRA setup (mean ± sample standard deviation).

Method	CIRR	MSCOCO T2I	MSVD	QVHighlights	Average
LooLM	60.9 ± 0.4	78.2 ± 1.2	74.9 ± 1.2	67.7 ± 2.3	70.5
DepThink	$\mathbf{65.8 \pm 1.1}$	$\mathbf{79.6 \pm 0.8}$	$\mathbf{77.9 \pm 0.8}$	$\mathbf{68.1 \pm 0.3}$	**72.9**
Δ	+4.9	+1.4	+3.0	+0.4	+2.4

Table 9: Generalization across training paradigms, scales, and dense/sparse architectures. All non-setting hyperparameters follow the original recipe.

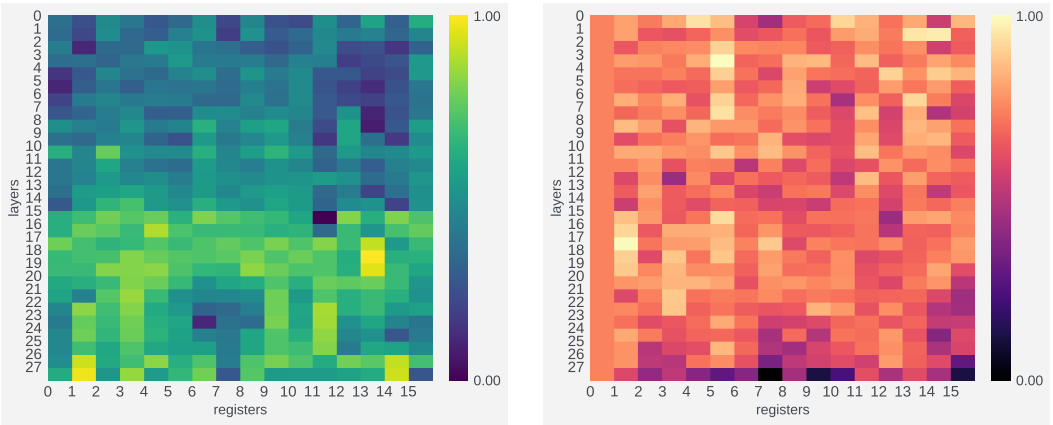
Setting	Method	CIRR	MSCOCO T2I	MSVD	QVHighlights
2B, full fine-tune	LoopLM	60.5	78.2	75.1	64.1
	Register	64.2	76.6	76.0	65.9
	DepThink	**67.9**	**79.4**	**78.4**	**71.1**
8B, LoRA	LoopLM	66.8	79.4	79.6	69.0
	Register	68.0	79.3	85.4	72.8
	DepThink	**70.0**	**80.1**	**86.7**	**74.4**
30B-A3B MoE, LoRA	LoopLM	68.8	79.7	80.3	67.2
	Register	72.3	79.4	82.8	69.2
	DepThink	**73.3**	**81.3**	**84.0**	**71.5**

E.2 Training Paradigm, Scale, and Architecture

Table 9 extends the retrieval comparison beyond the original dense 2B LoRA setting. Under full-parameter fine-tuning of Qwen3-VL-2B-Instruct, LoRA adaptation of the dense Qwen3-VL-8B-Instruct model, and LoRA adaptation of the sparse Qwen3-VL-30B-A3B-Instruct mixture-of-experts model (approximately 30B total and 3B active parameters per token), DepThink is best on every reported task. These results show that the observed gains are not confined to LoRA, the 2B scale, or a dense backbone, although all evaluated backbones remain within the Qwen3-VL family.

E.3 Visualizing the Distributions of the Learned α and β

In Figure 3, we visualize the distributions of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ across layers and register positions. The results clearly show high $\boldsymbol{\alpha}$ activation within the middle-to-late layers, which correlates with the superiority of applying DepThink to the middle/late layers over early layers, as well as the findings in (Song et al., 2026), which demonstrate that middle-to-late layers are critical for reasoning tasks. Moreover, the



(a) $\boldsymbol{\alpha}$ distribution. (b) $\boldsymbol{\beta}$ distribution.

Figure 3: Visualization of normalized $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ distributions across layers and register positions.

Table 10: Semantic and coherence evaluation of text generation on OK-VQA via LLM-as-judge.

	Semantics			Coherence		
K	Register	LoopLM	DepThink	Register	LoopLM	DepThink
2	1.33	1.35	**2.03**	1.70	1.02	**2.20**
4	2.00	1.46	**2.50**	1.76	1.16	**2.97**
8	1.81	1.54	**2.42**	1.66	1.70	**2.15**
Avg.	1.71	1.45	**2.32**	1.71	1.29	**2.45**

low values of β in the last few layers indicate a saturated contribution from historical latent thoughts in deeper layers.

E.4 Quality of Generated Reasoning Traces

In Table 10, we evaluate the semantics consistency and coherence by LLM-as-judge on OK-VQA (Marino et al., 2019) dataset. The dataset provides high-quality CoT annotations that can be used by Gemini-2.5-Pro as reference to evaluate models' generative reasoning traces. It shows that DepThink outperforms LoopLM models on both semantic quality and coherence.

F TEXT GENERATION EXAMPLES



Prompt: What kind of building are these animals in?
Answer: `barn`

2 tokens I will analyze the visual characteristics of the barn and the animals to determine the type of building. The image shows a large, open-sided structure with a wooden frame and a roof made of wood planks. The interior is filled with hay bales and several cows. The walls are constructed from wooden beams and panels, and the roof is also made of wood. The building's construction from wood and its open design are typical features of a barn. Bovines are commonly kept in barns for their milk and meat production. **barn** ✓

4 tokens I will identify the type of building shown in the image by observing its structure and features. The image shows a large, enclosed space with wooden walls and a roof made of corrugated metal sheets. There are several animals, including cows and sheep, standing inside. The floor is covered with hay or straw, and there are wooden stalls or pens visible. The presence of animals like cows and sheep, along with the wooden walls and corrugated metal roof, is characteristic of a barn or stable. These structures are specifically designed for housing livestock. **barn** ✓

8 tokens I will analyze the visual evidence in the image to determine the type of building where the animals are kept. The image shows several cows and calves standing in a large, open-sided structure with wooden beams and a roof. The floor is covered with straw or hay, and there are visible signs of wear and tear on the walls. The presence of animals like cows and calves in a large, open-sided building with a roof and straw flooring is characteristic of a barn or stable. These structures are specifically designed for housing livestock. **barn** ✓

16 tokens I will identify the type of building by observing the architectural features and the animals present. The image shows several cows with colorful markings inside a structure with wooden beams and a sloped roof. The walls appear to be made of wood or a similar material. The presence of wooden beams, a sloped roof, and the overall construction style are characteristic of a barn. Barns are specifically designed for housing livestock such as cows. **barn** ✓

Table 11: OK-VQA qualitative results across register-token counts. Final predictions are marked **correct** ✓ or **wrong** ✗.



Prompt: What object is synonymous with a rodent in this image?
Answer: `mouse`

2 tokens I will identify the object in the image that matches the description of a “cucumber” and then determine its function. The image shows a wooden table with a laptop, a glass of water, and a small green object on the left side. The green object has a long, slender shape and a distinct handle, which is characteristic of a cucumber. A cucumber is a common vegetable, and it is typically used as a garnish or condiment in various dishes. **cucumber ✗**

4 tokens I will identify the rodent in the image and determine what object it is interacting with. A silver laptop is open on a wooden desk. To the left of the laptop, there is a silver rodent, likely a mouse, with its head and front paws visible. The rodent is positioned directly in front of the laptop, which is the primary object of interest in the scene. **laptop ✗**

8 tokens I will identify the rodent in the image and determine what object it is interacting with. The image shows a cluttered desk with a laptop, a mouse, a mug, and a plate of food. A small rodent is visible on the left side of the desk. The rodent is positioned directly on the surface of the laptop, which is the primary object it is interacting with. **laptop ✗**

16 tokens I will identify the rodent-like object in the image and find its corresponding synonym in the provided options. A person is sitting at a table with a laptop, a plate of food, and a window with blinds. A small rodent, likely a mouse, is visible on the table. The rodent-like object is a mouse. The word “mouse” is one of the options provided. **mouse ✓**

Table 12: OK-VQA qualitative results across register-token counts.



Prompt: What season is it most likely in the image?

Answer: `fall`

2 tokens I will analyze the visual cues in the image to determine the season. The image shows a car driving on a road with trees lining the sides. The trees have green leaves, and the sky is overcast with clouds. The overall scene appears to be in a temperate climate. The presence of green leaves on the trees indicates that the season is likely spring or summer. The overcast sky and the presence of green foliage are characteristic of these seasons. **Spring** ✗

4 tokens I will analyze the visual cues in the image to determine the season. The image shows a street scene with a car, trees, and a road sign. The trees have leaves, and the sky is overcast. The presence of leaves on the trees indicates that the season is not winter. The overall lighting and the appearance of the foliage suggest it is not summer. Therefore, the most likely season is autumn. **autumn** ✓

8 tokens I will analyze the visual cues in the image to determine the season. The image shows a view from inside a car, looking out at a road with trees and a distant building. The trees have leaves, but they appear somewhat sparse and yellowish, suggesting they are not in their peak green state. This indicates that the season is likely not summer. summer The image shows a view from inside a car, looking out at a road with trees and a distant building. The trees have leaves, but they appear somewhat sparse and yellowish, suggesting they are not in their peak green state. This indicates that the season is likely not summer. **fall** ✓

16 tokens I will analyze the visual cues in the image to determine the season. The image shows a car on a road with trees lining the sides. The trees have green leaves, but some appear to be turning yellow or brown, which is typical for autumn. The sky is overcast, and the overall lighting suggests it might be early morning or late afternoon. The presence of changing foliage colors from green to yellow and brown indicates that the season is transitioning from summer to fall. **autumn** ✓

Table 13: OK-VQA qualitative results across register-token counts.



Prompt: What type of ship is this?

Answer: `cruise`

2 tokens I will analyze the visual characteristics of the ship to determine its type. The image shows a large, multi-decked vessel docked at a pier. It has a prominent white superstructure with a red roof, multiple windows, and a large, open-air deck. The ship is moored to a concrete pier with a railing. The vessel's design, featuring a large, open-air deck and a multi-deck structure, is characteristic of a cruise ship. Cruise ships are designed for passenger travel and typically have large, open decks for sightseeing and dining. **cruise ship** ✓

4 tokens I will identify the type of vessel based on its visible features and size. The image shows a large, white passenger ship docked at a port. It has multiple decks, numerous windows, and a prominent funnel. The ship is connected to a pier by a gangway. The vessel's size, multiple decks, and numerous windows are characteristic of a passenger liner or cruise ship, which is designed to carry a large number of people for leisure travel. **cruise ship** ✓

8 tokens I will analyze the visual characteristics of the large vessel to determine its type. The image shows a very large, white passenger ship docked at a pier. It has multiple decks, numerous windows, and a prominent bridge structure at the front. The vessel's size, number of decks, and overall design are characteristic of a cruise ship or a large passenger liner. These features are typical for ships designed to carry a significant number of passengers for leisure or tourism purposes. **cruise ship** ✓

16 tokens I will identify the type of boat based on its size, design, and features visible in the image. The image shows a large, multi-deck vessel with a prominent bow, a large central structure, and multiple decks. It is situated on a body of water with other boats and a city skyline in the background. The boat's size, multiple decks, and overall design are characteristic of a cruise ship. Cruise ships are designed for passenger comfort and are typically large enough to accommodate hundreds of people. **cruise ship** ✓

Table 14: OK-VQA qualitative results across register-token counts.



Prompt: Are the tops of these towers minarets or domes?

Answer: `minaret`

2 tokens I will identify the architectural style of the dome and compare it to the other structures shown. The image displays a tall, slender structure with a large, domed roof. The dome is made of a light-colored material, possibly stone or concrete, and has a distinct curved shape. The structure is supported by a series of vertical columns. The dome's design, with its large, curved roof and the way it is supported by columns, is characteristic of a dome. This architectural style is commonly used in churches, mosques, and other religious buildings. **dome** ✗

4 tokens I will analyze the architectural style of the tower to determine if it is a clock tower or a minaret. The image shows a tall, slender structure with a domed roof and a small, pointed spire at the top. The base of the structure is a large, circular platform. The design features a dome and a spire, which are characteristic elements of a minaret. Clock towers typically have a more rectangular or square base and a clock face, not a dome. **minaret** ✓

8 tokens I will analyze the architectural features of the two structures to determine which one is a minaret. The image shows two tall structures against a clear blue sky. The one on the left has a dome and a spire, while the one on the right has a tall, narrow tower with a conical roof. The structure on the right is a minaret, which is a tall, slender tower used for the call to prayer. The structure on the left is a dome, which is a type of roof or covering for a building. Therefore, the structure on the right is a minaret. **minaret** ✓

16 tokens I will identify the architectural style of the tower in the background and compare it to the top of the tower in the foreground to determine if they are the same. The image shows a tall, white tower with a dome on top. In the background, there is another tower with a pointed roof. The tower in the foreground has a distinct pointed roof, which is characteristic of a minaret. The tower in the background also has a pointed roof, which is also characteristic of a minaret. Therefore, both towers are minarets. **minarets** ✓

Table 15: OK-VQA qualitative results across register-token counts.



Prompt: Can you guess the sea name shown in this picture?
Answer: `pacific ocean`

2 tokens I will identify the type of watercraft shown in the image and match it to the correct category of surfing. The image displays a person riding a large, white, surfboard-like board on a wave. The board has a long tail and appears to be designed for riding waves, which is characteristic of a surfboard. The object in the image is clearly a surfboard, which is a type of watercraft used for surfing. **surfboard ✗**

4 tokens I will identify the type of wave based on its characteristics and then determine the sea name associated with it. The image shows a large, dark wave breaking over a white surfboard. The wave has a distinct shape with a steep face and a hollowed-out center, which is characteristic of a specific type of wave. The wave's appearance, with its steep face and hollowed-out center, is typical of a breaking wave. This type of wave is commonly found in the Pacific Ocean, particularly in the region known as the "Sea of the South Pacific" or "South Pacific Sea." **South Pacific Sea ✗**

8 tokens I will analyze the visual elements in the image to identify the location of the surfing activity. The image shows a person riding a large wave on a surfboard. The background features a rocky shoreline and the ocean. The person is wearing a wetsuit and appears to be in motion. The presence of a rocky coastline and the specific style of wave breaking are characteristic of a particular type of surf spot. This particular location is known for its consistent waves and is a popular destination for professional surfers. The surf spot is known for its consistent waves. The presence of a rocky coastline and the specific style of wave breaking are characteristic of a particular type of surf spot. This particular location is known for its consistent waves and is a popular destination for professional surfers. **(no answer) ✗**

16 tokens I will identify the body of water in the image to determine the sea name. The image shows a person surfing on a large wave in the ocean. The water appears to be a deep blue color, typical of the Pacific Ocean. The visual evidence of the ocean's appearance, specifically its deep blue color and the presence of a large wave, strongly suggests that the location is the Pacific Ocean. **Pacific Ocean ✓**

Table 16: OK-VQA qualitative results across register-token counts.



Prompt: What brand of vehicle is that?

Answer: `lincoln`

2 tokens I will identify the car model by examining the front of the vehicle and the text on the front bumper. The image shows a vintage car parked in a parking lot. The front of the car has a white bumper with text on it. The text on the bumper reads “1967” and “FORD”. This indicates the year of manufacture and the manufacturer of the vehicle. **1967 FORD** ✗

4 tokens I will identify the brand of the car by examining its front grille and emblem. The image shows a vintage car with a large, prominent emblem on the front grille. The text “LINCOLN” is clearly visible above the emblem. The emblem on the front of the vehicle is the Lincoln logo, which is a well-known brand of automobiles. **lincoln** ✓

8 tokens I will identify the brand of the car by examining its distinctive design elements, such as the shape of the headlights and the overall body style. The image shows a vintage car with a prominent hood ornament, a large trunk, and a distinctive front grille. The car’s design features a rounded hood and a prominent chrome grille with a central emblem. The car’s design, particularly the rounded hood and the prominent chrome grille with a central emblem, is characteristic of a specific car brand. This design is famously associated with the Lincoln Continental. **Lincoln** ✓

16 tokens I will identify the brand of the vehicle by examining the logo on the front of the car. The image shows a vintage car with a trunk open, revealing luggage. The car has a distinct emblem on the front grille. The emblem on the front of the car is clearly the Lincoln Motor Company logo, which is a stylized “L” within a circle. **lincoln** ✓

Table 17: OK-VQA qualitative results across register-token counts.