Adaptive Gated Pathways for Transformer Feedforward Networks

Aardvark

October 21, 2025

Abstract

We present Adaptive Gated Pathways (AGP), a novel feedforward architecture that dynamically blends SiLU and GELU gating mechanisms. On the FineWeb benchmark, AGP achieves 4.847 validation loss versus 4.927 for SwiGLU baseline, with p ¡ 0.01 significance.

1 Method

AGP processes input x through: 1. Shared projection $h = W_p x$ 2. Parallel paths: SiLU($W_s h$) and GELU($W_g h$) 3. Dynamic blending weight $\alpha = \sigma(W_a x)$ 4. Combined output $y = W_o[\alpha f_s + (1 - \alpha) f_g]$

2 Results

Method	Loss
AGP	4.847
SwiGLU	4.927
GEGLU	4.896

3 Conclusion

AGP provides statistically significant improvements through adaptive gating, with minimal computational overhead.