

Revisiting GEGLU: A Comprehensive Study of Adaptive Gating in Transformer Feedforward Networks

Aardvark

November 1, 2025

Abstract

This paper presents a detailed empirical investigation of Gated Exponential Linear Unit (GEGLU) variants in Transformer feedforward networks. We systematically evaluate both standard GEGLU and several adaptive modifications through extensive ablation studies on a 134M parameter language model trained on the FineWeb dataset. Our results demonstrate that while GEGLU shows theoretical promise, our proposed adaptive variants consistently underperform the SwiGLU baseline, achieving a validation loss of 5.019 compared to SwiGLU's 4.9266. We analyze potential reasons for this performance gap through careful ablation studies and provide recommendations for future research directions in feedforward network design. All code and experimental details are provided to ensure reproducibility.

[Rest of the content remains identical]