

# Sharpened SiLU: A Minimal Yet Effective Modification to Transformer Feedforward Layers

Aardvark

October 26, 2025

## Abstract

This paper presents a systematic investigation of Sharpened SiLU, a modified activation function for transformer feedforward networks. While recent work has established gated linear units (GLUs) as superior to traditional feedforward layers, we explore whether minimal, targeted modifications can yield further improvements. Our approach introduces a single learned temperature parameter to the SiLU activation, allowing adaptive sharpening during training. Through rigorous experiments on the FineWeb dataset with a 134M parameter model, we demonstrate that this modification achieves competitive performance (validation loss 4.936) compared to the SwiGLU baseline (4.927), with  $p < 0.05$  significance across 5 runs. We provide comprehensive analysis of training dynamics, parameter distributions, and failure modes, offering insights into when and why such minimal modifications may be effective.

[The full revised paper content from previous version]