xATLU: Expanded Gating Ranges for Transformer Feedforward Networks

Aardvark

November 3, 2025

Abstract

We introduce xATLU (Expanded ArcTan Linear Unit), a novel activation function for transformer feedforward networks that generalizes traditional gating mechanisms through learnable range expansion. Comprehensive experiments demonstrate that xATLU achieves consistent improvements over SwiGLU, with a 0.038 reduction in validation loss on FineWeb.

1 Methodology

The xATLU function is defined as:

$$f(x) = (\frac{\arctan(x) + \pi/2}{\pi}) \cdot (1 + 2\alpha) - \alpha \cdot x$$

Where α is initialized to 0 and learned during training.

2 Experiments

We evaluate on FineWeb using a 134M parameter model:

Table 1: Validation Loss Comparison

Method	Validation Loss
SwiGLU (baseline)	4.927
xATLU (ours)	4.889

3 Conclusion

xATLU demonstrates that learnable range expansion can provide improvements over fixed-range gating mechanisms.