

Compute Multiplier Estimates for Various Improvements

(CEG) Multiplier Estimate

10^1

10^0

--- No improvement (1.0x)

GELU vs ReLU

0.98

Cosine vs Inverse Sqrt Schedule

1.06

SwiGLU vs GeLU

1.09

Rotary vs Learned

1.19

Pre vs Post Norm

1.73

Adam vs SGD (bs=64)

3.10

Transformer vs LSTM

5.52

Improvement Type

