Activation Function Selection Guide

DEFAULT CHOICE

Start with ReLU

Fast, works well for most cases

Deep CNNs →

Leaky ReLU / PReLU

Prevents dead neurons in deep networks

Faster Convergence →

ELU

Slightly slower but better accuracy

Fully-Connected Deep Networks →

SELU

Eliminates need for Batch Normalization

Transformers / NLP →

GELU

Standard in BERT, GPT models

Limited Compute →

ReLU / Leaky ReLU

Fastest computational speed

Experimental / Research →

Swish / GELU

Potential accuracy gains

💡 **General rule: Modern smooth activations (GELU, Swish) often worth the extra computational cost**