ChimeraLM Architecture

Input — Tokenized DNA/RNA [B, L]

Hyena Operator — Convolution

Hidden States [B, L, 256]

Attention Pooling:
Linear(256→128)
GELU
Linear(128→1)
Softmax
Weighted Sum

Pooled Output [B, 256]

Classification:
Linear(256→512)
GELU + Dropout
Linear(512→512)
GELU + Dropout
ResidualBlock

Output — Linear(512→2) [B, 2]

[B, L] → [B, L, 256] → [B, L, 256] → [B, 256] → [B, 256] → [B, 512]

Legend:
B = batch size
L = sequence length