

Adaptive Lernrate mit Warm Up und Cosine Decay

