

Training Dynamics: Non-Monotonic Unfaithfulness Development  
Qwen3-0.6B, 20000 Documents

