**Eval Loss per Phase: Comparison of CoT Token Strategies
(Shaded region = ±1 std across 20 seeds)**

Legend:
- num_cot_tokens = 1
- num_cot_tokens = phase

X-axis: Phase
Y-axis: Eval Loss (MSE)