From one to three scaling laws

Intelligence (y-axis)

Compute (x-axis)

Pre-training scaling

Post-training scaling

Test-time scaling "long thinking