Training Loss Curve (Log-Log Scale)  $10^{-1}$ Exponential Loss  $10^{-3}$ 10<sup>0</sup> 10<sup>2</sup> 10<sup>3</sup> 10<sup>5</sup>  $10^{1}$  $10^{4}$ Training Steps (t)