The 1st Law of Information Capacity

$$\eta = \frac{D\left(H - L\right)}{N}$$

The 2nd Law of Information Capacity

$$\eta = \frac{E_0}{k_B N T \ln 2}$$

η: information capacity
N: parameter size (in bits)
L: average cross-entropy training loss
D: number of trained tokens
H: entropy of the entire dataset
E_0: minimum energy required to train
k_B: the Boltzmann constant
T: the Kelvin temperature of radiator