Causal Language Modeling

Dataset = MiniPile · Dataset = RedPajama · Dataset = LAMBADA · Dataset = Fineweb

$-\log p(x_t | x_{<t})$

Scaling Parameter (Pretraining FLOP)

Dataset
MiniPile
RedPajama
LAMBADA
Fineweb