

Test loss x energy consumption (lower is better)

Model size (MAE)

100M

600M

1B

8

16

32

64

128

GPUs

1e6

2.5

2.0

1.5

1.0

0.5

