

Training FLOPs Scaling for SOTA CV, NLP, and Speech Models

Training Compute (PFLOPs)

Transformer: 750x / 2 yrs
CV/NLP/Speech: 15x / 2 yrs
Moore's Law: 2x / 2 yrs

