

Training 4-layer GPT models on 4xA100 80GB SXM, 500GB RAM

