

Training 8-layer GPT models on 8xA100 80GB SXM, 900GB RAM

