Main experiment: training on OpenWebText