

Training

2024-02-20

Training

Subsections:

- [Model parallelism](#)
- [Performance](#)
- [Fault Tolerance](#)
- [Reproducibility](#)
- [Instabilities](#)
- [Checkpoints](#)
- [Training hyper-parameters and model initializations](#)
- [Tensor precision / Data types](#)
- [Emulate a multi-node setup using just a single node](#) - instructions on how to emulate a multi-node setup using just a single node - we use the `deepspeed` launcher here.
- [Re-train HF hub models from scratch using finetuning examples](#)

Tools:

- [printflock.py](#) - a tiny library that makes your `print` calls non-interleaved in a multi-gpu environment.
- [multi-gpu-non-interleaved-print.py](#) - a `flock`-based wrapper around `print` that prevents messages from getting interleaved when multiple processes print at the same time - which is the case with `torch.distributed` used with multiple-gpus.