

Comparison of Frameworks using ~7B Models for
Input & Output Length = 1024 on One A100 GPU

