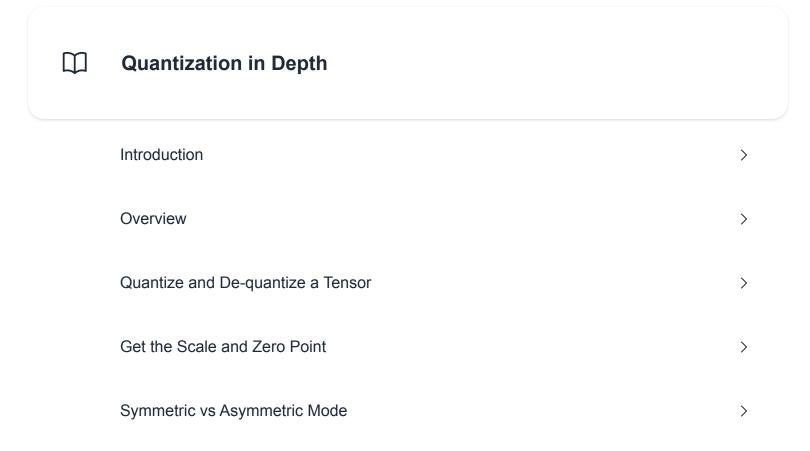


Lou Spironello, congratulations on completing Quantization in Depth!



Finer Granularity for more Precision	>
Per Channel Quantization	>
Per Group Quantization	>
Quantizing Weights & Activations for Inference	>
Custom Build an 8-Bit Quantizer	>
Replace PyTorch layers with Quantized Layers	>
Quantize any Open Source PyTorch Model	>
Load your Quantized Weights from HuggingFace Hub	>
Weights Packing	>
Packing 2-bit Weights	>
Unpacking 2-Bit Weights	>
Beyond Linear Quantization	>
Conclusion	>