Training Loss over Epochs

Legend:
- With Newton-Schulz
- Without Newton-Schulz
- Adaptive Newton-Schulz

X-axis: Epoch
Y-axis: L2 Loss