Training Loss Comparison Across Transformer Models