Vector Training Losses - Qwen2.5 14B
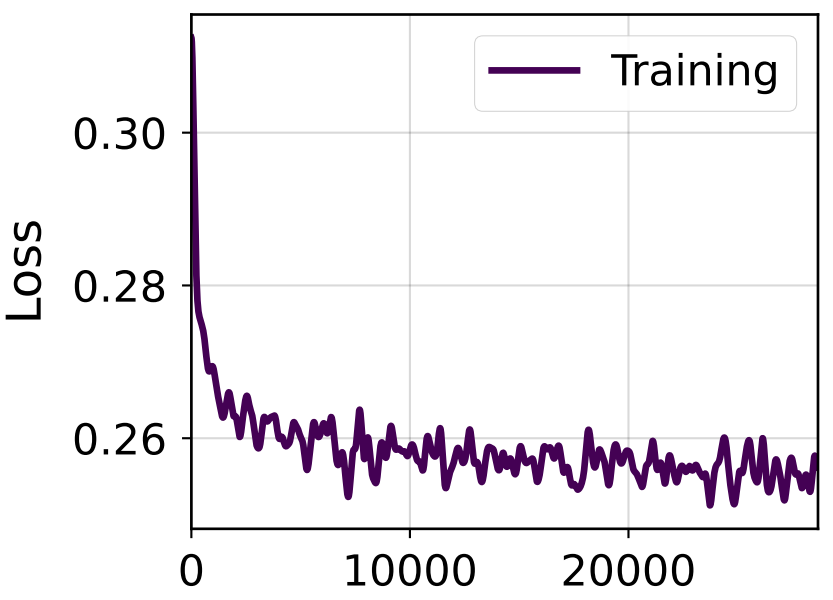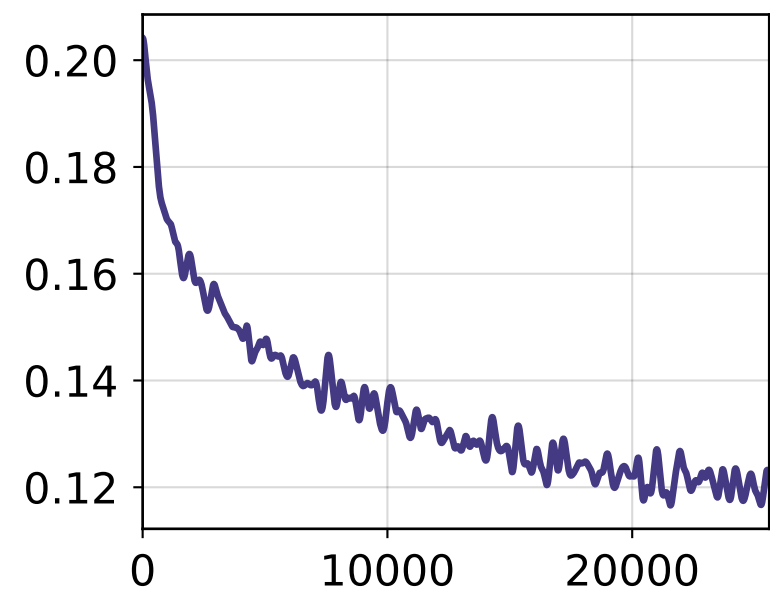
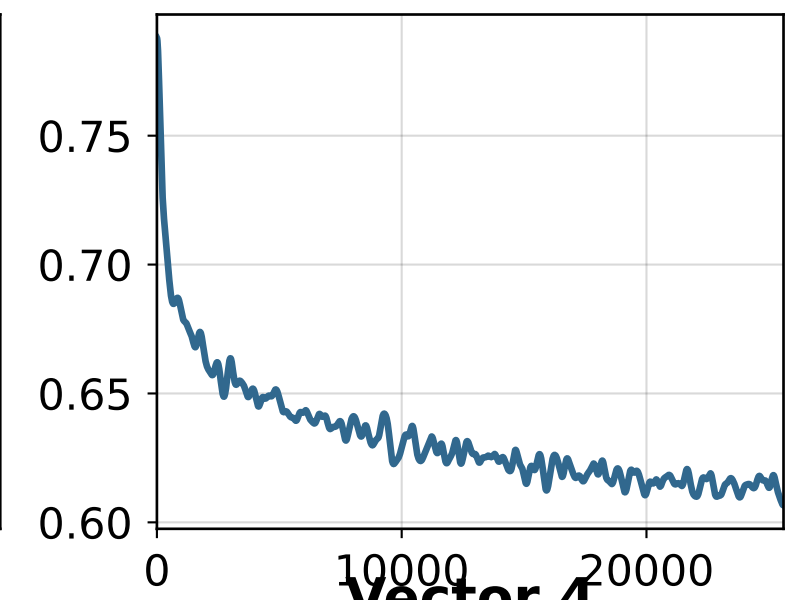Thinking model: Deepseek R1 Distill Qwen 14B | Layer 38 | Clusters 5