Vector Training Losses - Qwen2.5 32B

Thinking model: Deepseek R1 Distill Qwen 32B | Layer 27 | Clusters 15