Vector Training Losses - Llama 3.1 8B

Thinking model: Deepseek R1 Distill Llama 8B | Layer 6 | Clusters 15