

Vector Training Losses - Qwen2.5 Math 1.5B

Thinking model: Deepseek R1 Distill Qwen 1.5B | Layer 4 | Clusters 15

